STIC Search Report
Biotech-Chem Library

USPTO

STIC Database Tracking Number: 123552

TO: Celine Qian
Location: REM-2C70/2A89
Art Unit: 1636
Friday, June 04, 2004

Case Serial Number: 09/515363

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC SEARCH RESUL

FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➢ *I am an examiner in Workgroup:* _____   *Example: 1610*

➢ *Relevant prior art* **found**, *search results used as follows:*

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
  (journal articles, conference proceedings, new product announcements etc.)

➢ *Relevant prior art* **not found:**

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library   Remsen Bldg.

sira
Search and Information
Resources Administration

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.
Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**
Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

677N: Ed Hart.

12355 2 Access DB# _____

CRFE

SEARCH REQUEST FORM
Scientific and Technical Information Center

Requester's Full Name __Celine Qian__ Examiner #: __78710__ Date: __6/2/04__
Art Unit __1636__ Phone Number ~~30~~ 372-0777 Serial Number: __09/515363__
Mail Box and Bldg Room Location __2A89__ Results Format Preferred (circle): **PAPER** DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.
* *

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract

Title of Invention __Melanoma differentiation associated gene-5__

Inventors (please provide full names): __Paul Fisher et al.__

Earliest Priority Filing Date. __2/29/2000__

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number

__Please search for SEQ ID NO:1 & 2.__

- -

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June  2, 2004, 19:00:27 ; Search time 63 Seconds
                           (without alignments)
                           4597.003 Million cell updates/sec

Title:       US-09-515-363C-2
Perfect score: 5311
Sequence:    1 MSNGYSTDENFRYLISCFRA.........LPITFPNLDYSECLFSDED 1025

Scoring table:  BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:   1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
             Maximum Match 100%
             Listing first 45 summaries

Database :   A_Geneseq_29Jan04:*
             1: geneseqp1980s:*
             2: geneseqp1990s:*
             3: geneseqp2000s:*
             4: geneseqp2001s:*
             5: geneseqp2002s:*
             6: geneseqp2003as:*
             7: geneseqp2003bs:*
             8: geneseqp2004s:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	5311	100.0	5311	1025	4	AAE10155	Aae10155 Human mel
2	5285	99.5	5285	1025	5	AAM47798	Aam47798 Human RNA
3	5285	99.5	5285	1025	4	ADC31794	Adc31794 Human nov
4	4306	81.1	4306	838	4	AAE10165	Aae10165 RNA helic
5	2595.5	48.9	2456	558	4	AAU23090	Aau23090 Novel hum
6	2456	46.2	2456	468	6	ADA54899	Ada54899 Human pro
7	1860	35.0	1860	417	4	AAU23647	Aau23647 Novel hum
8	1839	34.6	1839	356	4	AAM40129	Aam40129 Human pol
9	1769	33.3	1769	348	4	AAM41915	Aam41915 Human pol
10	1769	33.3	1769	348	7	ADC33339	Adc33339 Human nov
11	1352	25.5	1352	678	5	ABP69672	Abp69672 Human pro
12	1344	25.3	1344	678	4	AAB93708	Aab93708 Human pol
13	1093	20.6	1093	304	4	AAU00296	Aau00296 Interfero
14	1093	20.6	1093	308	4	AAU00298	Aau00298 Interfero
15	861	16.2	861	267	4	AAU00297	Aau00297 Interfero
16	861	16.2	861	271	4	AAU00299	Aau00299 Interfero
17	808	15.2	808	448	4	AAU23654	Aau23654 Novel hum
18	783	14.7	783	166	4	ABB42219	Abb42219 Peptide #
19	783	14.7	783	166	4	AAM36026	Aam36026 Human bon
20	783	14.7	783	166	4	AAM75915	Aam75915 Human bon
21	783	14.7	783	166	4	AAM63104	Aam63104 Human bra
22	782	14.7	782	154	4	AAG74314	Aag74314 Human col
23	782	14.7	782	154	5	ABP41112	Abp41112 Human ova
24	666.5	12.5	666.5	357	3	AAB42981	Aab42981 Human ORF
25	563.5	10.6	563.5	447	5	ABP41171	Abp41171 Human ova
26	558.5	10.5	447			AAB95485	Aab95485 Human pro
27	521	9.8	312			AAU23099	Aau23099 Novel hum
28	488	9.2	92			AAM14138	Aam14138 Peptide #
29	488	9.2	92			ABB33083	Abb33083 Peptide #
30	488	9.2	92			AAM26544	Aam26544 Peptide #
31	488	9.2	92			ABB27911	Abb27911 Human pep
32	488	9.2	92			ABB18551	Abb18551 Protein #
33	488	9.2	92			AAM66268	Aam66268 Human bon
34	488	9.2	92			AAM53880	Aam53880 Human bra
35	488	9.2	92			ABG47932	Abg47932 Human liv
36	488	9.2	92			AAM01876	Aam01876 Peptide #
37	488	9.2	92			AEG35915	Abg35915 Human pep
38	453	8.5	752			AAB96107	Aab96107 Putative
39	406	7.6	1909			AAB31166	Aab31166 Amino aci
40	395.5	7.4	764			AAB62028	Aab62028 Recombina
41	357	6.7	71			ABB42510	Abb42510 Peptide #
42	357	6.7	71			AAM36323	Aam36323 Peptide #
43	357	6.7	71			AAM76213	Aam76213 Human bon
44	332.5	6.3	1383			AEB63586	Abb63586 Drosophil
45	332	6.3	1924			AAE09768	Aae09768 Human dic

ALIGNMENTS

```
RESULT 1
AAE10155
ID   AAE10155 standard; protein; 1025 AA.
XX
AC   AAE10155;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Human melanoma differentiation associated (Mda)-5 protein.
XX
KW   Human; melanoma differentiation associated gene; Mda-5; interferon; IFN;
KW   RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;
KW   neuroblastoma; astrocytoma; glioblastoma; multiforme; cancer; cervical;
KW   breast; colon; prostate; osteosarcoma; chrondosarcoma; systemic toxicity;
KW   central nervous system; cytostatic; apoptosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          125..174
FT                   /note="Reveals sequence homology with other CARD
FT                   proteins"
FT   Domain          331..336
FT                   /note="ATPase A motif"
FT   Domain          443..446
FT                   /note="ATPase B motif"
FT   Region          722..823
FT                   /note="Reveals significant homology to RNA helicase C-
FT                   terminal conserved domain"
XX
PN   WO200164707-A1.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US006960.
XX
PR   29-FEB-2000; 2000US-00515363.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Fisher PB, Kang D, Gopalkrishnan RV;
XX
DR   WPI; 2001-565494/63.
DR   N-PSDB; AAD17203.
XX
PT   Nucleic acid sequences encoding a Melanoma Differentiation Associated
PT   Gene useful for cancer cell growth suppresion, apoptosis and anti-viral
PT   activity.
```

XX
PS   Claim 21; Page 18-19; 152pp; English.
XX
CC   The present invention relates to an isolated nucleic acid encoding a
CC   melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
CC   contains a caspase recruitment domain (CARD) and a RNA helicase motif.
CC   Mda-5 is a novel interferon (IFN) inducible gene with structural
CC   similarities to RNA helicases and CARD motif containing proteins. Mda-5
CC   is induced during terminal differentiation in human melanoma cells
CC   treated with the combination of recombinant fibroblast IFN and the
CC   antileukaemic compound mezerein (MEZ). Mda-5 is useful for identifying
CC   compounds that may induce its expression. Mda-5 is useful for treating
CC   cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
CC   multiforme, cervical cancer, breast cancer, colon cancer, prostate
CC   cancer, osteosarcoma, chrondosarcoma and especially where the cancer is a
CC   cancer of the central nervous system and apoptosis. The Mda-5 promoter
CC   exhibits melanocyte tissue specificity and minimises systemic toxicity.
CCC The present sequence is human Mda-5 protein
XX
SQ   Sequence 1025 AA;

Query Match   100.0%; Score 5311; DB 4; Length 1025;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 1025; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYILTFLPAEVKEQIQRTVATSGNMQAVE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYILTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy   121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy   181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy   241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300

Qy   301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVI 360

Qy   361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy   421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYIMQKLKNNRLKKENKPVIP 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYIMQKLKNNRLKKENKPVIP 480

Qy   481 LPQIILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 LPQIILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540

Qy   541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAAKGNRKERVC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAAKGNRKERVC 600

Qy   601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFPAVIEDDSDEGGDDEYCDGDEDE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFPAVIEDDSDEGGDDEYCDGDEDE 660

Qy   661 DDLKKPLKLDETDRFLMTLFFENNKOMLKRLAEBNPEYNEKLTKLRNTIMEQYTRTEESAR 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAEBNPEYNEKLTKLRNTIMEQYTRTEESAR 720

Qy   721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKRRT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRT 780

Qy   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840

Qy   841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900

Qy   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960

Qy   961 ICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKQNSTKKQYKKWVELPITFPNLDYSECCL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKQNSTKKQYKKWVELPITFPNLDYSECCL 1020

Qy   1021 FSDED 1025
              |||||
Db   1021 FSDED 1025

RESULT 2
AAM47798
ID   AAM47798 standard; protein; 1025 AA.
XX
AC   AAM47798;
XX
DT   05-MAR-2002 (first entry)
XX
DE   Human RNA helicase RH116.
XX
KW   Human; RH116; RNA helicase; cytostatic; virucide; anti-HIV;
KW   immunosuppressive; immunostimulatory; antirheumatic; antiarthritic;
KW   antiarteriosclerotic; osteopathic; antidiabetic; hepatotropic;
KW   antiinflammatory; cancer; infection; HIV; hepatitis; genetic disease;
KW   autoimmune disease; graft rejection; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200185955-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-FR001441.
XX
PR   11-MAY-2000; 2000FR-00006030.
XX
PA   (ISTA-) ISTAC.
PA   (INSP) INST PASTEUR LILLE.
XX
PI   Bahr G, Cocude C, Capron A;
XX
DR   WPI; 2002-082898/11.
DR   N-PSDB; ABA04908.
XX
PT   New polypeptide, useful for treating and diagnosing cancer or
PT   inflammation, and drug screening, comprises a human polynucleotide
PT   homologous to RNA helicase.
XX
PS   Claim 1; Page 89-93; 114pp; French.
XX
CC   The present sequence is the protein sequence for human RH116. RH116 is a
CC   116kDa protein and has homology to RNA helicases (DEXH box). RH116 and
CC   its coding sequence are useful for treating cancer; acute or chronic
CC   infections (especially by HIV or hepatitis B or C); inherited genetic
CC   diseases; (auto)immune diseases (particularly rheumatism, arthritis,
CC   arteriosclerosis, osteoporosis and diabetes, but many others listed) and
CC   to prevent graft rejection. RH116 and its coding sequence are also useful
CC   for inducing, or increasing, the immune response to a vaccine

Db   961 ICKCCQAWGTMVHKGLDLPCLKIRNFVVFKNNSTKKQYKKWVELPITFPNLDYSECCL 1020

Qy   1021 FSDED 1025
          |||||
Db   1021 FSDED 1025

```
RESULT 3
ADC31794
ID   ADC31794 standard; protein; 1025 AA.
XX
AC   ADC31794;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human novel polypeptide sequence, SEQ ID NO:1876.
XX
KW   Human; diagnostic; drug screening; forensics; gene mapping;
KW   biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW   neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW   ulcers; osteoporosis; autoimmune disease; cancer;
KW   molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW   neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW   antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO2003029271-A2.
XX
PD   10-APR-2003.
XX
PF   24-SEP-2002; 2002WO-US030474.
XX
PR   24-SEP-2001; 2001US-0324631P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang TY,  Zhang J,  Ren F,  Xue AJ,  Zhao QA,  Wang J,  Wehrman T;
PI   Zhou P,  Ghosh M,  Wang D,  Ma Y,  Asundi V,  Wang Z,  Weng G;
PI   Haley-Vicente D,  Drmanac RT;
XX
DR   WPI; 2003-371981/35.
DR   N-PSDB; ADC30823.
XX
PT   New polynucleotide and polypeptide useful for diagnosing, preventing or
PT   treating conditions such as neurodegenerative diseases, anemias, platelet
PT   disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT   cancer.
XX
PS   Claim 20; SEQ ID NO 1876; 1185pp; English.
XX
CC   The invention relates to 971 novel human cDNA sequences (ADC29919-
CC   ADC30889)and the polypeptides they encode (ADC30890-ADC31860). The
CC   invention also relates to nucleic acid sequences over 99% identical with
CC   the novel human cDNAs. The invention additionally encompasses expression
CC   vectors and host cells comprising a nucleic acid of the invention; the
CC   recombinant production of a polypeptide of the invention; an antibody
CC   against a polypeptide of the invention; a method of detecting
CC   polynucleotides or polypeptides of the invention; and methods of
CC   identifying a compound which binds to a polypeptide of the invention. The
CC   invention further discloses methods of preventing, treating or
CC   ameliorating a medical condition; kits comprising polynucleotide probes
CC   and/or monoclonal antibodies for carrying out the methods of the
CC   invention; methods for the identification of compounds that modulate the
CC   expression or activity of the polynucleotide and/or polypeptide; and 767
CC   contig sequences corresponding to the cDNA sequences of the invention
CC   (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC   -ADC33394). The nucleic acids and polypeptides of the invention are
CC   useful in diagnostics, drug screening, forensics, gene mapping, in the
CC   identification of mutations responsible for genetic disorders or other
CC   traits, for assessing biodiversity, and in producing many other types of
CC   data and products dependent on DNA and amino acid sequences. They are
```

XX
SQ   Sequence 1025 AA;

Query Match   99.5%; Score 5285; DB 5; Length 1025;
Best Local Similarity   99.5%; Pred. No. 0;
  Matches 1020; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

```
Qy     1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy    61 LLLSTLEKGVWHIGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LLLSTLEKGVWHIGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy   121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy   181 FLNVLRQTGNNELVQELTGSDSCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 FLNVLRQTGNNELVQELTGSDSCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy   241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGDSDSGTMGSDSDEENVAARA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGDSDSGTMGSDSDEENVAARA 300

Qy   301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVI 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVI 360

Qy   361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy   421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNRLKKENKPVIP 480
         ||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db   421 LLNLENGEDAGVQLSDFSFIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480

Qy   481 LPQILGLTASPGVGGATKQAKAREEHILKLCANLDAFTIKTVKENLDQLKNQIOEPCKKFA 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 LPQILGLTASPGVGGATKQAKAREEHILKLCANLDAFTIKTVKENLDQLKNQIOEPCKKFA 540

Qy   541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKGNRKERVC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||| || |||
Db   541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKESVC 600

Qy   601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AEHLRKYNKALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660

Qy   661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720

Qy   721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780

Qy   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840

Qy   841 IEHETVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKOMKTKRNIAKHYKNNP 900
         || :|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 IERETVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKOMKTKRNIAKHYKNNP 900

Qy   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960

Qy   961 ICKCCQAWGTMVHKGLDLPCLKIRNFVVFKNNSTKKQYKKWVELPITFPNLDYSECCL 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
CC  also used for treating diseases such as Parkinson's disease, Alzheimer's
CC  disease and other neurodegenerative diseases, anaemia, platelet
CC  disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC  cancer. The nucleic acids may also be used as hybridisation probes or
CC  primers, and in the recombinant production of a protein. The polypeptides
CC  are also useful in generating antibodies, as molecular weight markers,
CC  and as food supplements. The present sequence represents a specifically
CC  claimed human polypeptide sequence of the invention. Note: The sequence
CC  data for this patent did not form part of the printed specification, but
CC  was obtained in electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 1025 AA;

Query Match      99.5%;   Score 5285;   DB 7;   Length 1025;
Best Local Similarity   99.5%;   Pred. No. 0;
Matches 1020;   Conservative   2;   Mismatches   2;   Indels   3;   Gaps   0;

Qy     1  MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYITFLPAEVKEQIQRTVATSGNMQAVE    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYITFLPAEVKEQIQRTVATSGNMQAVE    60

Qy    61  LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL   120

Qy   121  LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENGNESGVRELLKRIVQKENWFSA   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENGNESGVRELLKRIVQKENWFSA   180

Qy   181  FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM   240

Qy   241  ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA   300

Qy   301  SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI   360

Qy   361  VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS   420

Qy   421  LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNRLKKENKPVIP   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  LLNLENGEDAGVQLSDFSFIIIDECHHTNKEAVYNNIMRHYLMQKLKNRLKKENKPVIP   480

Qy   481  LPQIGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  LPQIGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA   540

Qy   541  IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC   600
          ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db   541  IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKSGNRKRSVC   600

Qy   601  AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE   660
          |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  AEHLRKYNKALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE   660

Qy   661  DDLKKPLKLDETDRFLMTLFPENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR   720

Qy   721  GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSSEFKPMTQNEQKEVISKFRT   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSSEFKPMTQNEQKEVISKFRT   780

Qy   781  GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV   840

Qy   841  IEHBTVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP   900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  IERETVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP   900

Qy   901  SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI   960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI   960

Qy   961  ICKCGQAWGTMVVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCL  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  ICKCGQAWGTMVVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCL  1020

Qy  1021  FSDED  1025
          |||||
Db  1021  FSDED  1025


RESULT 4
AAE10165
ID  AAE10165 standard; protein; 838 AA.
XX
AC  AAE10165;
XX
DT  29-NOV-2001  (first entry)
XX
DE  RNA helicase conserved motif of human Mda-5 protein.
XX
KW  Human; melanoma differentiation associated gene; Mda-5; interferon; IFN;
KW  RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;
KW  neuroblastoma; astrocytoma; glioblastoma; multiforme; cancer; cervical;
KW  breast; colon; prostate; osteosarcoma; chrondosarcoma; systemic toxicity;
KW  central nervous system; cytostatic; apoptosis.
XX
OS  Homo sapiens.
XX
PN  WO20016707-A1.
XX
PD  07-SEP-2001.
XX
PF  28-FEB-2001; 2001WO-US006960.
XX
PR  29-FEB-2000; 2000US-00515363.
XX
PA  (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI  Fisher PB,  Kang D,  Gopalkrishnan RV;
XX
DR  WPI; 2001-565494/63.
XX
PT  Nucleic acid sequences encoding a Melanoma Differentiation Associated
PT  Gene useful for cancer cell growth suppression, apoptosis and anti-viral
PT  activity.
XX
PS  Example 1; Fig 1D; 152pp; English.
XX
CC  The present invention relates to an isolated nucleic acid encoding a
CC  melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
CC  contains a caspase recruitment domain (CARD) and a RNA helicase motif.
CC  Mda-5 is a novel interferon (IFN) inducible gene with structural
CC  similarities to RNA helicases and CARD motif containing proteins. Mda-5
CC  is induced during terminal differentiation in human melanoma cells
CC  treated with the combination of recombinant fibroblast IFN and the
CC  antileukaemic compound mezerein (ME2). Mda-5 is useful for identifying
CC  compounds that may induce its expresion. Mda-5 is useful for treating
CC  cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
CC  multiforme, cervical cancer,breast cancer, colon cancer, prostate
CC  cancer, osteosarcoma, chrondosarcoma and especially where the cancer is a
CC  cancer of the central nervous system and apoptosis. The Mda-5 promoter
CC  exhibits melanocyte tissue specificity and minimises systemic toxicity.
CC  The present sequence is RNA helicase conserved motif of human Mda-5
CC  protein
XX
```

```
SQ   Sequence 838 AA;

Query Match             81.1%;   Score 4306;   DB 4;   Length 838;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches   838;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy     61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 60

Qy    121 LQPTIVDKLLIVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LQPTIVDKLLIVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 120

Qy    181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 180

Qy    241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ENNSSESSFADSSVVSSSDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 240

Qy    301 SPEPEIQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 SPEPEIQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 300

Qy    361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 360

Qy    421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 420

Qy    481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 480

Qy    541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 540

Qy    601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 600

Qy    661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 660

Qy    721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 720

Qy    781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 780

Qy    841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKN 898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKN 838
```

```
RESULT 5
AAU23090
ID   AAU23090 standard; protein; 558 AA.
XX
AC   AAU23090;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human enzyme polypeptide #176.
XX
```

```
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW   nephrotropic; anticoagulant.
XX
OS   Homo sapiens.
XX
PN   WO20015301-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001239.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184464P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
```

```
PR    14-SEP-2000;    2000US-0233064P.
PR    14-SEP-2000;    2000US-0233065P.
PR    21-SEP-2000;    2000US-0234223P.
PR    21-SEP-2000;    2000US-0234274P.
PR    25-SEP-2000;    2000US-0234997P.
PR    25-SEP-2000;    2000US-0234998P.
PR    26-SEP-2000;    2000US-0235484P.
PR    27-SEP-2000;    2000US-0235834P.
PR    27-SEP-2000;    2000US-0235836P.
PR    29-SEP-2000;    2000US-0236327P.
PR    29-SEP-2000;    2000US-0236367P.
PR    29-SEP-2000;    2000US-0236368P.
PR    29-SEP-2000;    2000US-0236369P.
PR    29-SEP-2000;    2000US-0236370P.
PR    02-OCT-2000;    2000US-0239937P.
PR    13-OCT-2000;    2000US-0239935P.
PR    13-OCT-2000;    2000US-0240960P.
PR    20-OCT-2000;    2000US-0241221P.
PR    20-OCT-2000;    2000US-0241785P.
PR    20-OCT-2000;    2000US-0241786P.
PR    20-OCT-2000;    2000US-0241787P.
PR    20-OCT-2000;    2000US-0241808P.
PR    20-OCT-2000;    2000US-0241809P.
PR    20-OCT-2000;    2000US-0241826P.
PR    01-NOV-2000;    2000US-0244617P.
PR    08-NOV-2000;    2000US-0246474P.
PR    08-NOV-2000;    2000US-0246475P.
PR    08-NOV-2000;    2000US-0246476P.
PR    08-NOV-2000;    2000US-0246477P.
PR    08-NOV-2000;    2000US-0246478P.
PR    08-NOV-2000;    2000US-0246523P.
PR    08-NOV-2000;    2000US-0246524P.
PR    08-NOV-2000;    2000US-0246525P.
PR    08-NOV-2000;    2000US-0246526P.
PR    08-NOV-2000;    2000US-0246527P.
PR    08-NOV-2000;    2000US-0246528P.
PR    08-NOV-2000;    2000US-0246532P.
PR    08-NOV-2000;    2000US-0246609P.
PR    08-NOV-2000;    2000US-0246610P.
PR    08-NOV-2000;    2000US-0246611P.
PR    08-NOV-2000;    2000US-0246613P.
PR    17-NOV-2000;    2000US-0249207P.
PR    17-NOV-2000;    2000US-0249208P.
PR    17-NOV-2000;    2000US-0249209P.
PR    17-NOV-2000;    2000US-0249210P.
PR    17-NOV-2000;    2000US-0249211P.
PR    17-NOV-2000;    2000US-0249212P.
PR    17-NOV-2000;    2000US-0249213P.
PR    17-NOV-2000;    2000US-0249214P.
PR    17-NOV-2000;    2000US-0249215P.
PR    17-NOV-2000;    2000US-0249216P.
PR    17-NOV-2000;    2000US-0249217P.
PR    17-NOV-2000;    2000US-0249218P.
PR    17-NOV-2000;    2000US-0249244P.
PR    17-NOV-2000;    2000US-0249245P.
PR    17-NOV-2000;    2000US-0249264P.
PR    17-NOV-2000;    2000US-0249265P.
PR    17-NOV-2000;    2000US-0249297P.
PR    17-NOV-2000;    2000US-0249299P.
PR    17-NOV-2000;    2000US-0249300P.
PR    01-DEC-2000;    2000US-0250160P.
PR    01-DEC-2000;    2000US-0250391P.
PR    05-DEC-2000;    2000US-0251030P.
PR    05-DEC-2000;    2000US-0251988P.
PR    05-DEC-2000;    2000US-0256719P.
PR    06-DEC-2000;    2000US-0251479P.
PR    08-DEC-2000;    2000US-0251856P.
PR    08-DEC-2000;    2000US-0251868P.
```

```
PR    08-DEC-2000;    2000US-0251869P.
PR    08-DEC-2000;    2000US-0251989P.
PR    08-DEC-2000;    2000US-0251990P.
PR    11-DEC-2000;    2000US-0254097P.
PR    05-JAN-2001;    2000US-0259678P.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,   Barash SC,   Ruben SM;
XX
DR    WPI; 2001-465566/50.
-DR   N-PSDB; AAS40960.
XX
PT    Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT    treating neural, immune system, muscular, reproductive, pulmonary,
PT    cardiovascular, renal, proliferative disorders and cancerous diseases.
XX
PS    Claim 11; SEQ ID NO 1086; 1180pp; English.
XX
CC    The present invention relates to the isolation of novel human enzyme
CC    polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC    encoding them. The enzyme polypeptides of the invention may comprise the
CC    functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC    isomerases or ligases. The sequences of the invention are useful in the
CC    diagnosis, treatment, prevention and/or prognosis of a wide range of
CC    disorders including hyperproliferative disorders (e.g. cancer),
CC    immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC    arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC    disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC    cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC    (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC    infectious disorders (e.g. Influenza). The polynucleotides of the
CC    invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC    the novel human enzyme polypeptides of the invention. Note: The sequence
CC    data for this patent did not form part of the printed specification, but
CC    was obtained in electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 558 AA;

Query Match             48.9%;   Score 2595.5;  DB 4;  Length 558;
Best Local Similarity   94.6%;   Pred. No. 5.7e-192;
Matches  510;  Conservative   10;  Mismatches   16;  Indels    3;  Gaps    2;

Qy    294  ENVAARASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKA  353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ENVAARASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKA   60

Qy    354  SEPGKVIVLVNKVLLVBQLFRKEFQPFLKKWYRVIGLSGDTQLKISPEVVKSCDIIIST  413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  SEPGKVIVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISPEVVKSCDIIIST  120

Qy    414  AQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKK  473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKK  180

Qy    474  ENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQ  533
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQ  240

Qy    534  EPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKG  593
              ||||||||||||||||||||||||||||||||||||||||||||||| :|::||:
Db    241  EPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEG  300

Qy    594  NRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEY  653
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  NRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEY  360

Qy    654  CDGDEDEDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYT  713
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  CDGDEDEDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYT  420

Qy    714  RTEBSARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNBQKE  773
```

Db   421 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
            RTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKE 480

Qy   774 VISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAM-VQARGRARADESTYV 831
                |||||||||||| :::   .|::   :|::.: ||||||||||||||||
Db   481 VISKFRTGKINLLIAPQWQKKVWILK--NVTLXVMVSSPMKYHGPARGRARADESTYV 537

RESULT 6
ADA54899
ID   ADA54899 standard; protein; 468 AA.
XX
AC   ADA54899;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Human protein, SEQ ID 2467.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI   Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   N-PSDB; ADA53260.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 14; SEQ ID NO 2467; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 468 AA;

Query Match   46.2%; Score 2456; DB 6; Length 468;
Best Local Similarity   99.1%; Pred. No. 2.8e-181;
Matches 464; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   558 MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTI 617
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKERVCAEHLRKYNEALQINDTI 60

Qy   618 RMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM 677
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 RMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM 120

Qy   678 TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW 737
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW 180

Qy   738 ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLD 797
                |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db   181 ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGRINLLIATTVAEEGLD 240

Qy   798 IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKA 857
                ||||||||||||||||||||||||||||||||||||||||||||||||||.|||||||
Db   241 IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIERETVNDFREKMMYKA 300

Qy   858 IHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG 917
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 IHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG 360

Qy   918 EDIHVIEKMHVNMTPEFFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMMVHKGL 977
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 EDIHVIEKMHVNMTPEFKELYIVRENKTLQKKCADYQINGEIICKCGQAWGTMMVHKGL 420

Qy   978 DLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 468

RESULT 7
AAU23647
ID   AAU23647 standard; protein; 417 AA.
XX
AC   AAU23647;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Novel human enzyme polypeptide #733.
XX
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW   nephrotropic; anticoagulant.
XX
OS   Homo sapiens.
XX
PN   WO200155301-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001239.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0199123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225447P.

```
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   14-AUG-2000;  2000US-0225759P.
PR   18-AUG-2000;  2000US-0226279P.
PR   22-AUG-2000;  2000US-0226681P.
PR   22-AUG-2000;  2000US-0226868P.
PR   22-AUG-2000;  2000US-0271182P.
PR   23-AUG-2000;  2000US-0277009P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   06-SEP-2000;  2000US-0230437P.
PR   06-SEP-2000;  2000US-0230438P.
PR   08-SEP-2000;  2000US-0231242P.
PR   08-SEP-2000;  2000US-0231243P.
PR   08-SEP-2000;  2000US-0231244P.
PR   08-SEP-2000;  2000US-0231413P.
PR   08-SEP-2000;  2000US-0231414P.
PR   08-SEP-2000;  2000US-0232080P.
PR   08-SEP-2000;  2000US-0232081P.
PR   12-SEP-2000;  2000US-0231968P.
PR   14-SEP-2000;  2000US-0232397P.
PR   14-SEP-2000;  2000US-0232398P.
PR   14-SEP-2000;  2000US-0232399P.
PR   14-SEP-2000;  2000US-0232400P.
PR   14-SEP-2000;  2000US-0232401P.
PR   14-SEP-2000;  2000US-0233063P.
PR   14-SEP-2000;  2000US-0233064P.
PR   14-SEP-2000;  2000US-0233065P.
PR   21-SEP-2000;  2000US-0244223P.
PR   21-SEP-2000;  2000US-0234274P.
PR   25-SEP-2000;  2000US-0244998P.
PR   26-SEP-2000;  2000US-0235484P.
PR   27-SEP-2000;  2000US-0235834P.
PR   27-SEP-2000;  2000US-0235836P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236367P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0239935P.
PR   13-OCT-2000;  2000US-0299337P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241221P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241786P.
PR   20-OCT-2000;  2000US-0241787P.
PR   20-OCT-2000;  2000US-0241808P.
PR   20-OCT-2000;  2000US-0241809P.
PR   20-OCT-2000;  2000US-0241826P.
PR   01-NOV-2000;  2000US-0244617P.
PR   08-NOV-2000;  2000US-0246474P.
PR   08-NOV-2000;  2000US-0246475P.
PR   08-NOV-2000;  2000US-0246476P.
PR   08-NOV-2000;  2000US-0246477P.
PR   08-NOV-2000;  2000US-0246478P.
PR   08-NOV-2000;  2000US-0245523P.
PR   08-NOV-2000;  2000US-0245524P.
PR   08-NOV-2000;  2000US-0245525P.
PR   08-NOV-2000;  2000US-0245526P.
PR   08-NOV-2000;  2000US-0245527P.
PR   08-NOV-2000;  2000US-0246528P.
PR   08-NOV-2000;  2000US-0246532P.
PR   08-NOV-2000;  2000US-0246609P.
PR   08-NOV-2000;  2000US-0246610P.
PR   08-NOV-2000;  2000US-0246611P.
PR   08-NOV-2000;  2000US-0246613P.
PR   17-NOV-2000;  2000US-0249207P.
PR   17-NOV-2000;  2000US-0249208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251988P.
PR   05-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465566/50.
DR   N-PSDB; AAS41517.
XX
PT   Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT   treating neural, immune system, muscular, reproductive, pulmonary,
PT   cardiovascular, renal, proliferative disorders and cancerous diseases.
XX
PS   Claim 11; SEQ ID NO 1643; 1180pp; English.
XX
CC   The present invention relates to the isolation of novel human enzyme
CC   polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC   encoding them. The enzyme polypeptides of the invention may comprise the
CC   functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC   isomerases or ligases. The sequences of the invention are useful in the
CC   diagnosis, treatment, prevention and/or prognosis of a wide range of
CC   disorders including hyperproliferative disorders (e.g. cancer),
CC   immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC   arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC   disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC   cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC   (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC   infectious disorders (e.g. Influenza). The polynucleotides of the
CC   invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC   the novel human enzyme polypeptides of the invention. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 417 AA;

Query Match             35.0%;  Score 1860;  DB 4;  Length 417;
Best Local Similarity   90.8%;  Pred. No. 3.6e-135;
```

DR   N-PSDB: AAI59285.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Example 5; SEQ ID NO 3274; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 356 AA;

Query Match   34.6%;   Score 1839;   DB 4;   Length 356;
Best Local Similarity   99.4%;   Pred. No. 1.2e-133;
Matches 350; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   674 RFLMTLFFENNKKLKRLAENPEYENEKLTKLRNTIMEQYRTEESARGIIFTKTRQSAYA 733
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 RFLMTLFFENNKKLKRLAENPEYENEKLTKLRNTIMEQYRTEESARGIIFTKTRQSAYA 64

Qy   734 LSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRTGKINLLIATTVAE 793
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 LSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRTGKINLLIATTVAE 124

Qy   794 EGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKM 853
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 EGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKM 184

Qy   854 MYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVL 913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 MYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVL 244

Qy   914 ACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEIICKCCGQAWGTMMV 973
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 ACSGEDIHVIEKMHHVNMTPEFKELYIVRENKTLQKKCADYQINGEIICKCCGQAWGTMMV 304

Qy   974 HKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 HKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 356

RESULT 9
AAM41915
ID   AAM41915 standard; protein; 348 AA.
XX
AC   AAM41915;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 6846.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX

Matches 365; Conservative   13; Mismatches   20; Indels   4; Gaps   3;

Qy   442 IDECHHTNKEAVVNNIMRHYLMQKLKNRLKKENKPVIPLPQILGLTASPGVGGATKQAK 501
          : |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8 VDECHHTNKEAVVNNIMRHYLMQKLKNRLKKENKPVIPLPQILGLTASPGVGGATKQAK 67

Qy   502 AEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRI 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 AEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRI 127

Qy   562 QTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTIRMID 621
          |||||||||||||||||||||||||| : ||||||||||||||||||||||||||||||
Db   128 QTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKERVCAEHLRKYNEALQINDTIRMID 187

Qy   622 AYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFF 681
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   188 AYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFF 247

Qy   682 ENNKMLKRLAENPEYENEKLTKLRNTIMEQYRTEESARGIIFTKTRQSAYALSQWITEN 741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248 ENNKMLKRLAENPEYENEKLTKLRNTIMEQYRTEESARGIIFTKTRQSAYALSQWITEN 307

Qy   742 EKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRTGKINLLIATTVAEEGLDIKEC 801
          |||||||||||||||||||||||||||||||||||||||||| :: :: ||||
Db   308 EKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRTGKINLLIAPQWQKKVWILK-- 365

Qy   802 NIVIRYGLVTNEIAM-VQARGRARADESTYVLVAHSGSGVIE 842
          |: : : :|::: :|::: ||||| : :|||
Db   366 NVTLXSVMVSSPMKYHGPARGRARADESTYVWLL-QWSGVID 406

RESULT 8
AAM40129
ID   AAM40129 standard; protein; 356 AA.
XX
AC   AAM40129;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 3274.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000;   2000US-00488725.
PR   25-APR-2000;   2000US-00552317.
PR   20-JUN-2000;   2000US-00598042.
PR   19-JUL-2000;   2000US-00620312.
PR   03-AUG-2000;   2000US-00653450.
PR   14-SEP-2000;   2000US-00662191.
PR   19-OCT-2000;   2000US-00693036.
PR   29-NOV-2000;   2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI   Zhou P, Goodrich R, Drmanac RT;
XX
DR   WPI; 2001-442253/47.

```
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            303 PCLKIRNFVVFKONSTKKQYKKWVELPITFPNLDYSECCLFSDED 34B
```

```
RESULT 10
ADC33339
ID   ADC33339 standard; protein; 348 AA.
XX
AC   ADC33339;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Human novel contig-encoded polypeptide sequence, SEQ ID NO:3421.
XX
KW   Human; diagnostic; drug screening; forensics; gene mapping;
KW   biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW   neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW   ulcers; osteoporosis; autoimmune disease; cancer;
KW   molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW   neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW   antiulcer; osteopathic; immunosuppressive; antinflammatory; cytostatic;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO2003029271-A2.
XX
PD   10-APR-2003.
XX
PF   24-SEP-2002; 2002WO-US030474.
XX
PR   24-SEP-2001; 2001US-0324631P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang TY,  Zhang J,  Ren F,  Xue AJ,  Zhao QA,  Wang J,  Wehrman T;
PI   Zhou P,  Ghosh M,  Wang D,  Ma Y,  Asundi V,  Wang Z,  Weng G;
PI   Haley-Vicente D,  Drmanac RT;
XX
DR   WPI; 2003-371981/35.
DR   N-PSDB; ADC32572.
XX
PT   New polynucleotide and polypeptide useful for diagnosing, preventing or
PT   treating conditions such as neurodegenerative diseases, anemias, platelet
PT   disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT   cancer.
XX
PS   Example 2; SEQ ID NO 3421; 1185pp; English;
XX
CC   The invention relates to 971 novel human cDNA sequences (ADC29919-
CC   ADC30889)and the polypeptides they encode (ADC30890-ADC31860). The
CC   invention also relates to nucleic acid sequences over 99% identical with
CC   the novel human cDNAs. The invention additionally encompasses expression
CC   vectors and host cells comprising a nucleic acid of the invention; the
CC   recombinant production of a polypeptide of the invention; an antibody
CC   against a polypeptide of the invention; a method of detecting
CC   polynucleotides or polypeptides of the invention; and methods of
CC   identifying a compound which binds to a polypeptide of the invention. The
CC   invention further discloses methods of preventing, treating or
CC   ameliorating a medical condition; kits comprising polynucleotide probes
CC   and/or monoclonal antibodies for carrying out the methods of the
CC   invention; methods for the identification of compounds that modulate the
CC   expression or activity of the polynucleotide and/or polypeptide; and
CC   contig sequences corresponding to the cDNA sequences of the invention
CC   (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC   -ADC33394). The nucleic acids and polypeptides of the invention are
CC   useful in diagnostics, drug screening, forensics, gene mapping, in the
CC   identification of mutations responsible for genetic disorders or other
CC   traits, for assessing biodiversity, and in producing many other types of
CC   data and products dependent on DNA and amino acid sequences. They are
CC   also used for treating diseases such as Parkinson's disease, Alzheimer's
CC   disease and other neurodegenerative diseases, anaemia, platelet
CC   disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
```

```
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;  99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang Y,  Zhao QA;
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI61071.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Example 2; SEQ ID NO 6846; 1007pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 348 AA;
```

```
Query Match       33.3%;  Score 1769;  DB 4;  Length 348;
Best Local Similarity  97.4%;  Pred. No. 3.1e-128;
Matches  337;  Conservative   2;  Mismatches   2;  Indels   7;  Gaps   0;

Qy    680  FFENNKMLKRLAENPEYENEKLITKLRNTIMEQYRTEBSARGIIFTKIRQSAYALSQWIT 739
                : ||  || |||||||||| ||||||||||||||||||||||||||||||
Db      3  FLKTIKXLNRLAEHPXYENEKLITKLRNTIMEQYRTEBSARGIIFTKIRQSAYALSQWIT 62

Qy    740  ENEKFAEVGKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIK 799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63  ENEKFAEVGKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIK 122

Qy    800  ECNIVIRYGLVTNEIAMVQARGRARADESTYLVAHSGSGVIBHETVNDFREKMMYKAIH 859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123  ECNIVIRYGLVTNEIAMVQARGRARADESTYLVLVAHSGSGVIBHETVNDFREKMYKAIH 182

Qy    860  CVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGED 919
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183  CVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGED 242

Qy    920  IHVIEKMHHVNMTPFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMVHKGLDL 979
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243  IHVIEKMHHVNMTPFKELYIVRENKTLQKKCADYQINGEIICKCGQAWGTMVHKGLDL 302

Qy    980  PCLKIRNFVVFKONSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
```

```
CC   cancer. The nucleic acids may also be used as hybridisation probes or
CC   primers, and in the recombinant production of a protein. The polypeptides
CC   are also useful in generating antibodies, as molecular weight markers,
CC   and as food supplements. The present sequence represents a human contig-
CC   encoded polypeptide sequence used in an example of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 348 AA;

Query Match            33.3%;  Score 1769;  DB 7;  Length 348;
Best Local Similarity  97.4%;  Pred. No. 3.1e-128;
Matches 337;  Conservative  2;  Mismatches  7;  Indels  0;  Gaps  0;

Qy  680 FFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWIT 739
             | ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    3 FLKTIKLNRLAEHPXYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWIT  62

Qy  740 ENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIK 799
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIK 122

Qy  800 ECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIH 859
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 ECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIH 182

Qy  860 CVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGED 919
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183 CVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGED 242

Qy  920 IHVIEKMHHVNMTPEFKELYIVRENKALQKCADYQINGEIICKCGQAWGTMVHKGLDL 979
        |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  243 IHVIEKMHHVNMTPEFKELYIVRENKTLQKKCADYQINGEIICKCGQAWGTMVHKGLDL 302

Qy  980 PCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
        |||||||||||||||||||||||||||||||||||||||||||||
Db  303 PCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED  348

RESULT 11
ABP69672
ID   ABP69672 standard; protein; 678 AA.
XX
AC   ABP69672;
XX
DT   20-JAN-2003 (first entry)
XX
DE   Human polypeptide SEQ ID NO 1719.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US005095.
XX
PR   05-MAR-2001; 2001US-00799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI   Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   N-PSDB; ABZ11889.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT   or coagulation disorders.
XX
PS   Claim 9; SEQ ID NO 1719; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences (ABZ11119-
CC   ABZ12066) or their mature protein coding portion, active domain coding
CC   protein or complementary sequences. The polynucleotides are useful for
CC   identifying expressed genes or for physical mapping of human genome. The
CC   encoded polypeptides (ABP69902-ABP69949) are useful as molecular weight
CC   markers, as a food supplement, for generating antibodies, in medical
CC   imaging, screening and diagnostic assays and for treating cell-
CC   proliferative disorders (cancer) neurodegenerative diseases (Parkinson's
CC   or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC   diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC   platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC   or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC   arthritis, etc. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 678 AA;

Query Match            25.5%;  Score 1352;  DB 5;  Length 678;
Best Local Similarity  40.7%;  Pred. No. 1.8e-95;
Matches 292;  Conservative 136;  Mismatches 237;  Indels 52;  Gaps 12;

Qy  306 LGLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK 365
          :||: ||| | ||||||||||| |||| |||||  |||  ||||   :  | || |||:
Db    1 MELRSYQWEVIMPALEGKNIIIWLPTGAGKTRAAAYVAKRHLE---TVDGAKVVLVNR  56

Qy  366 VLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
         ||:| :|  :  | || ||  | |||: | ||:|:| ||| ||| :|:  |::  :: ||
Db   57 VHLVTQ-HGEEFRRMLDGRWTVTTLSGDMGPRAGFGHLARCHDLLICTAELLQMALTSPE 115

Qy  426 NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNRLKKENKPVIPLPQIL 485
         || | :||:||:|||:|:||||: : |:|||:| ||:|:|::|| |         ||| |
Db  116 --BEEHVELTVFSLIVVDBCHHTHKDTYVNIVMSQYLELKLQRAQ--------PLPQVL 164

Qy  486 GLTASPGVGGATKQAKAEEHILKLCANLDAFTITKTVKENLDQLKNQIOEPCKKFAIADAT 545
        ||||||| |||: |:: |||||||||||| | |  : || ||:||:|  :  ||  |:|:
Db  165 GLTASPGTGGASKLDGAINHVLQLCANLDTWCIMSPQNCCPQLQEHSQQPCKQYNLCHRR 224

Qy  546 REDPFKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL 604
         :|||| |||::|  || ||  :  |: || || ||:|| ||: :::| ::  |||  :||
Db  225 SQDPFGDLLKKLMDQIHDHLEMPELSRKFGTQMYEQQVVKLSEAAALAGLQBQRVYALHL 284

Qy  605 RKYNEALQINDTIRMIDAYTHLETFYNEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLK 664
        |:||:|| |:||:|:|| : :| :||:|:        : :  : ::       C----- 322
Db  285 RRYNDALLIHDTVRAVDLALAALQDFYHREHVTKTQIL----------C---------- 322

Qy  665 KPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIF 724
              :   |:  ||:||:|||||||:||||||||||||||||||||||||||||::|:||||
Db  323 ------AERRLLALFDDRKNELAHLATHGP-ENPKLEMLEKILQRQFS-SSNSPRGIIF 373

Qy  725 TKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKIN 784
        |:|||||::|| |:        |::|| |||||||:|||  ||||:: ::||::|  || :
Db  374 TRTRQSAHSLLLWLQQQGLQTVDIRAQLLIGAGNSSQSTHMTQRDQQEVIQKFQDGTLN 433

Qy  785 LLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHE 844
        || |: |||||||| ||: |||||:|||| :||||||||||:| :||  |:  ||: : :
Db  434 LLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRARADQSVYAFVATEGSRELKRE 493

Qy  845 TVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLIT 904
         ||: ::|  || :: ::     :: : :   ::  :  ::                  :
```

XX
SQ   Sequence 678 AA;

Query Match   25.3%;   Score 1344; DB 4; Length 678;
Best Local Similarity   40.6%;   Pred. No. 7.5e-95;
Matches 291; Conservative 136; Mismatches 238; Indels 52; Gaps 12;

Qy   306 LQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK 365
         ::|||:|::::::|: |||||||||||||| | : || :: |:::|||| |:
Db   1 MELRSYQWEVIMPALEGKNIIIWLPTGAGKTRAAYVAKRHLE---TVDGAKVVLVNR 56

Qy   366 VLLVEQLFRKEFQPPLKKWYRVIGLSGDTQLKISPEVVKSCDIIISTAQILENSLLNLE 425
         |:||| ||:: | |:| |: |||| || |:: :|| |: | ||| :| |
Db   57 VHLVTQ-HGEFRRMLDGRWTVTTLSGDMGPRAGFGHLARCHDLLICTAELLQMALTSPE 115

Qy   426 NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIL 485
         : : || : :| ::|| ||| | | :|:::: |: | ---PLPQVL 164
Db   116 --EBEHVELTVFSLIVVDECHHTHKDTVYNVIMSQYLELKIQRAQ---------PLPQVL 164

Qy   486 GLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT 545
         ||||||||||:|:|| | ::|: :|||:| || || ::: ::: |
Db   165 GLTASPGTGGASKLDGAINHVLQLCANLDTWCIMSPQNCCPQLQEHSQQPCKQYNLCHRR 224

Qy   546 REDPFKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL 604
         ::||||| | :: ::| | |: ||| ::|:|| :||: ::|: ::||| |:|::|||
Db   225 SQDPFGDLLKKLMDQIHDHLEMPELSRKFGTQMYEQQVVKLSERAAALAGLQEQRVYALHL 284

Qy   605 RKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLK 664
           ||::| | |: || : :: | : :   -C----- 322
Db   285 RRYNDALLIHDTVRAVDALAALQDFYHREHVTKTQIL----------------C------ 322

Qy   665 KPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIF 724
         | : ||::| |: ::::| : : :|: ::| |: | : :|: ||| ||| |||
Db   323 ------ABRRLLALFDDRKNELAHLATHGP-ENPKLEMLEKILQRFS-SSNSPRGIIF 373

Qy   725 TKTRQSAYALSQWITENEKPAEVGVKAHHLIGAGHSSBFKPMTQNEQKEVISKFRTGKIN 784
         |:||||:| |||:|||:|:||: | : ||:||| ::: ||| :|: |:|:| ||:|:|:|
Db   374 TRTRQSAHSLLLWLQQQGLQTVDIRAQLLIGAGNSSQSTHMTQRDQQEVIQKFQDGTLN 433

Qy   785 LLIATTVAEBGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHE 844
         || |||:|||| |||:|:| ||||:|||:|||||||||| :| ||::|:|||||||:|
Db   434 LLVATSVAEBGLDIPHCNVVVRYGLLTNEISMVQARGRAWADQSVTAFVATEGSREBKRE 493

Qy   845 TVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLIT 904
         | |:| ||: |::|:|:|||||:|||:|:||||:|| :||:||:||| |||||:|||
Db   494 LINEALETLMEQAVAAVQKMDQAEYQAKIRDLQQAALTKRAAQAQRENQRQQFPVEHVQ 553

Qy   905 FLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELY-IVRENKALQKKCADYQINGEIICK 963
         |||:||||:|::|||||::||::||:|:||||||| :|::: |:|||: |::|||:: |
Db   554 LLCINCMVAVGHGSDLRKVEGTHHVNVNPNFSNYYNVSRDPVVINKVFKDWKPGGVISCR 613

Qy   964 -CGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKVVELPITFPNLDYSECC 1019
          :|||||||: |||: || |: ::|:::| :: | |:|:|: | | | :| ||
Db   614 NCGEVWGLQMIYKSVKLPVLKVRS--MLLETPQGRIQAKKWSRVFPSVPDFDFLQHC 668

RESULT 12
AAB93708
ID   AAB93708 standard; protein; 678 AA.
XX
AC   AAB93708;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:13299.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999; 99JP-00248036.
PR   27-AUG-1999; 99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 13299; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are also useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH3166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention

Db   494 LINEALETLMEQAVAAVQKMDQAEYQAKIRDLQQAALTKRAAQAQRENQRQQFPVEHVQ 553

Qy   905 FLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELY-IVRENKALQKKCADYQINGEIICK 963
         |||:||||:|::|||||::||::||:|:||||||| :|::: |:|||: |::|||:: |
Db   554 LLCINCMVAVGHGSDLRKVEGTHHVNVNPNFSNYYNVSRDPVVINKVFKDWKPGGVISCR 613

Qy   964 -CGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKVVELPITFPNLDYSECC 1019
          :|||||||: |||: || |: ::|:::| :: | |:|:|: | | | :| ||
Db   614 NCGEVWGLQMIYKSVKLPVLKVRS--MLLETPQGRIQAKKWSRVFPSVPDFDFLQHC 668

RESULT 13
AAU00296
ID   AAU00296 standard; protein; 304 AA.
XX
AC   AAU00296;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Interferon induced polypeptide, IFN4.
XX
KW   Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
KW   immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
KW   graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
KW   human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
KW   haematologic disease; chronic neutropenia; myocardial infarction;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
KW   amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN4.

XX
OS   Homo sapiens.
XX
PN   WO20011820B-A2.
XX
PD   15-MAR-2001.
XX
PF   08-SEP-2000; 2000WO-US024704.
XX
PR   08-SEP-1999;   99US-0152921P.
PR   20-OCT-1999;   99US-0160575P.
PR   20-JAN-2000;   2000US-0177104P.
PR   07-SEP-2000;   2000US-0655633.
XX
PA   (CURA-) CURAGEN CORP.
PA   (BIOJ) BIOGEN INC.
XX
PI   Peyman JA,   Da Silva A,   Hochman P,   Hsu A;
XX
DR   WPI; 2001-235201/24.
DR   N-PSDB; AAS01149.
XX
PT   New interferon induced polypeptides and polynucleotides, useful for the
PT   diagnosis, prevention and treatment of immunological, cell proliferative
PT   disorders, such as lupus erythematosus, cancer, stroke and Alzheimer's
PT   disease.
XX
PS   Claim 1; Page 29-32; 134pp; English.
XX
CC   The sequence represents interferon induced polypeptide, IFN4. IFN nucleic
CC   acids and polypeptides are useful for treating or preventing a pathology
CC   associated with IFN polypeptide in a human. They are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC   acids, polypeptides and antibodies are useful for diagnosis, prevention
CC   or treatment of variety of immunological and cell proliferative
CC   disorders, such as autoimmune diseases e.g. lupus erythematosus,
CC   immunological diseases such as acquired immunodeficiency syndrome
CC   (AIDS), graft rejection, viral infections including hepatitis and human
CC   immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
CC   haematologic diseases such as aplastic anaemia and chronic neutropenia
CC   and cancer. In addition they are also useful for treating or diagnosing
CC   various disorders associated with cell death, including myocardial
CC   infarction, stroke, neurological diseases including Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC   atrophy. IFN nucleic acids and polypeptides are also useful for
CC   identifying interferon-like proteins and interferon agonists, for
CC   screening drugs and compounds which inhibit or enhance IFN activity or
CC   function and as targets for the identification of small molecules that
CC   are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC   cell or tumour cell growth in mammals, including humans
XX
SQ   Sequence 304 AA;

Query Match   20.6%;   Score 1093; DB 4;   Length 304;
Best Local Similarity   100.0%;   Pred. No. 6e-76;
Matches   216;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   41   EVKEQIQRTVATSGNMQAVELLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELT 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87   EVKEQIQRTVATSGNMQAVELLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELT 146

Qy 101   DLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEBELLTIEDRNRIAAAENNGN 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 147   DLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEBELLTIEDRNRIAAAENNGN 206

Qy 161   ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQV 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 207   ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQV 266

Qy 221   EEQLLSTTVQPNLEKEVWGMENNSSESSFADSSVVS 256
              |||||||||||||||||||||||||||||||||||
Db 267   EEQLLSTTVQPNLEKEVWGMENNSSESSFADSSVVS 302

RESULT 14
AAU00298
ID   AAU00298 standard; protein; 308 AA.
XX
AC   AAU00298;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Interferon induced polypeptide, IFN6.
XX
KW   Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
KW   immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
KW   graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
KW   human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
KW   haematologic disease; chronic neutropenia; myocardial infarction;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
KW   amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN6.
XX
OS   Homo sapiens.
XX
PN   WO20011820B-A2.
XX
PD   15-MAR-2001.
XX
PF   08-SEP-2000; 2000WO-US024704.
XX
PR   08-SEP-1999;   99US-0152921P.
PR   20-OCT-1999;   99US-0160575P.
PR   20-JAN-2000;   2000US-0177104P.
PR   07-SEP-2000;   2000US-0655633.
XX
PA   (CURA-) CURAGEN CORP.
PA   (BIOJ) BIOGEN INC.
XX
PI   Peyman JA,   Da Silva A,   Hochman P,   Hsu A;
XX
DR   WPI; 2001-235201/24.
DR   N-PSDB; AAS01151.
XX
PT   New interferon induced polypeptides and polynucleotides, useful for the
PT   diagnosis, prevention and treatment of immunological, cell proliferative
PT   disorders, such as lupus erythematosus, cancer, stroke and Alzheimer's
PT   disease.
XX
PS   Claim 1; Page 40-42; 134pp; English.
XX
CC   The sequence represents interferon induced polypeptide, IFN6. IFN nucleic
CC   acids and polypeptides are useful for treating or preventing a pathology
CC   associated with IFN polypeptide in a human. They are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC   acids, polypeptides and antibodies are useful for diagnosis, prevention
CC   or treatment of variety of immunological and cell proliferative
CC   disorders, such as autoimmune diseases e.g. lupus erythematosus,
CC   immunodeficiency diseases such as acquired immunodeficiency syndrome
CC   (AIDS), graft rejection, viral infections including hepatitis and human
CC   immunodeficiency virus (HIV), immune-mediated glomerulonephritis and human
CC   haematologic diseases such as aplastic anaemia and chronic neutropenia
CC   and cancer. In addition they are also useful for treating or diagnosing
CC   various disorders associated with cell death, including myocardial
CC   infarction, stroke, neurological diseases including Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC   atrophy. IFN nucleic acids and polypeptides are also useful for
CC   identifying interferon-like proteins and interferon agonists, for
CC   screening drugs and compounds which inhibit or enhance IFN activity or
CC   function and as targets for the identification of small molecules that
CC   are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC   cell or tumour cell growth in mammals, including humans
XX
SQ   Sequence 308 AA;

```
Query Match        20.6%;  Score 1093; DB 4;  Length 308;
Best Local Similarity  100.0%;  Pred. No. 6.1e-76;
Matches  216; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy    41 EVKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91 EVKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELT 150

Qy   101 DLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEEBLLTIEDRNRIAAAENNGN 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 DLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEEBLLTIEDRNRIAAAENNGN 210

Qy   161 ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQV 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQV 270

Qy   221 EEQLLSTTVQPNLEKEVWGMENNSESSFADSSVVS 256
         ||||||||||||||||||||||||||||||||||
Db   271 EEQLLSTTVQPNLEKEVWGMENNSESSFADSSVVS 306
```

RESULT 15
AAU00297
ID   AAU00297 standard; protein; 267 AA.
XX
AC   AAU00297;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Interferon induced polypeptide, IFN5.
XX
KW   Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
KW   immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
KW   graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
KW   human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
KW   haematologic disease; chronic neutropenia; myocardial infarction;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
KW   amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN5.
XX
OS   Homo sapiens.
XX
PN   WO200118208-A2.
XX
PD   15-MAR-2001.
XX
PF   08-SEP-2000; 2000WO-US024704.
XX
PR   08-SEP-1999; 99US-0152921P.
PR   20-OCT-1999; 99US-0160575P.
PR   20-JAN-2000; 2000US-0177104P.
PR   07-SEP-2000; 2000US-0655633.
XX
PA   (CURA-) CURAGEN CORP.
PA   (BIOJ) BIOGEN INC.
XX
PI   Peyman JA, Da Silva A, Hochman P, Hsu A;
XX
DR   WPI; 2001-235201/24.
DR   N-PSDB; AAS01150.
XX
PT   New interferon induced polypeptides and polynucleotides, useful for the
PT   diagnosis, prevention and treatment of immunological, cell proliferative
PT   disorders, such as lupus erythematosus, cancer, stroke and Alzheimer's
PT   disease.
XX
PS   Claim 1; Page 33-35; 134pp; English.
XX
CC   The sequence represents interferon induced polypeptide, IFN5. IFN nucleic
CC   acids and polypeptides are useful for treating or preventing a pathology
CC   associated with IFN polypeptide in a human. They are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC   acids, polypeptides and antibodies are useful for diagnosis, prevention
CC   or treatment of variety of immunological and cell proliferative
CC   disorders, such as autoimmune diseases e.g. lupus erythematosus,
CC   immunodeficiency diseases such as acquired immunodeficiency syndrome
CC   (AIDS), graft rejection, viral infections including hepatitis and human
CC   immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
CC   haematologic diseases such as aplastic anaemia and chronic neutropenia
CC   and cancer. In addition they are also useful for treating or diagnosing
CC   various disorders associated with cell death, including myocardial
CC   infarction, stroke, neurological diseases including Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC   atrophy. IFN nucleic acids and polypeptides are also useful for
CC   identifying interferon-like proteins and interferon agonists, for
CC   screening drugs and compounds which inhibit or enhance IFN activity or
CC   function and as targets for the identification of small molecules that
CC   are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC   cell or tumour cell growth in mammals, including humans
XX
SQ   Sequence 267 AA;

```
Query Match        16.2%;  Score 861; DB 4;  Length 267;
Best Local Similarity  96.6%;  Pred. No. 4.7e-58;
Matches  170; Conservative   2; Mismatches   4; Indels   0; Gaps   0;

Qy    41 EVKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    87 EVKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELT 146

Qy   101 DLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEEBLLTIEDRNRIAAAENNGN 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   147 DLPSPSFENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEEBLLTIEDRNRIAAAENNGN 206

Qy   161 ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVD 216
         |||||||||||||||||  |  ::
Db   207 ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAGICNFTEED 262
```

Search completed: June 2, 2004, 19:10:14
Job time : 67 secs

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   June 2, 2004, 19:08:58 ; Search time 23 Seconds
                                   (without alignments)
                           2300.723 Million cell updates/sec

Title:          US-09-515-363C-2
Perfect score:  5311
Sequence:       1 MSNGYSTDENFRYLISCFRA........LPITFPNLDYSECCLFSDED 1025

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:     389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   Issued Patents AA:*
             1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
             2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
             3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
             4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
             5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
             6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	286	5.4	503	4	US-09-408-020-66	Sequence 66, Appl
2	284	5.3	502	4	US-09-408-020-34	Sequence 34, Appl
3	217.5	4.1	2662	4	US-09-595-684B-31	Sequence 31, Appl
4	202	3.8	3878	1	US-09-914-259-11	Sequence 11, Appl
5	194.5	3.7	2482	1	US-09-328-254-6	Sequence 6, Appli
6	193.5	3.6	661	4	US-09-107-532A-3677	Sequence 3677, Ap
7	191.5	3.6	3248	1	US-08-353-700-1	Sequence 1, Appli
8	191.5	3.6	3248	5	PCT-US95-16216-1	Sequence 1, Appli
9	188.5	3.5	1388	2	US-08-685-576-4	Sequence 4, Appli
10	188.5	3.5	1388	4	US-09-976-594-296	Sequence 296, App
11	187.5	3.5	2125	4	US-09-919-172-29	Sequence 29, Appl
12	186.5	3.5	1786	4	US-08-973-462-8	Sequence 8, Appli
13	182.5	3.4	2954	4	US-09-150-867-1	Sequence 1, Appli
14	179	3.4	666	4	US-09-134-001C-5465	Sequence 5465, Ap
15	174	3.3	1388	2	US-08-685-576-1	Sequence 1, Appli
16	173	3.3	1211	4	US-09-134-001C-4820	Sequence 4820, Ap
17	171.5	3.2	976	3	US-09-104-324B-4	Sequence 4, Appli
18	171.5	3.2	2285	2	US-09-308-375-2	Sequence 2, Appli
19	167.5	3.2	1219	4	US-09-107-532A-6020	Sequence 6020, Ap
20	167	3.1	677	4	US-09-328-352-4365	Sequence 4365, Ap
21	167	3.1	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
22	166.5	3.1	1886	5	US-08-938-105-3	Sequence 3, Appli
23	164	3.1	1354	3	US-08-685-187-1	Sequence 1, Appli
24	163	3.1	1939	4	US-09-310-187A-1	Sequence 1, Appli
25	159.5	3.0	956	4	US-09-134-001C-4452	Sequence 4452, Ap
26	159	3.0	442	4	US-09-489-039A-11770	Sequence 11770, A
27	158	3.0	781	2	US-08-675-631-3	Sequence 3, Appli
28	158	3.0	781	4	US-08-359-316A-3	Sequence 3, Appli
29	158	3.0	781	4	US-09-248-776-3	Sequence 3, Appli
30	157.5	3.0	590	4	US-09-543-681A-7931	Sequence 7931, Ap
31	156.5	2.9	1581	4	US-09-866-108A-15754	Sequence 15754, A
32	156	2.9	407	4	US-09-399-081A-11	Sequence 11, Appl
33	156	2.9	961	4	US-09-914-259-66	Sequence 66, Appl
34	155	2.9	454	4	US-09-543-681A-7717	Sequence 7717, Ap
35	155	2.9	1261	3	US-09-208-742-4	Sequence 4, Appli
36	155	2.9	1261	4	US-09-332-295-2	Sequence 2, Appli
37	155	2.9	1261	4	US-09-709-979-2	Sequence 2, Appli
38	155	2.9	1261	4	US-10-147-268-2	Sequence 2, Appli
39	154.5	2.9	875	4	US-09-107-532A-4537	Sequence 4537, Ap
40	154.5	2.9	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
41	153.5	2.9	394	4	US-09-134-000C-4332	Sequence 4332, Ap
42	152	2.9	407	3	US-08-989-370-6	Sequence 6, Appli
43	152	2.9	407	4	US-09-398-169-6	Sequence 6, Appli
44	152	2.9	1111	4	US-09-914-259-28	Sequence 28, Appl
45	151.5	2.9	709	4	US-09-702-705-335	Sequence 335, App

ALIGNMENTS

```
RESULT 1
US-09-408-020-66
; Sequence 66, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCORP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 66
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-66

Query Match             5.4%;  Score 286;  DB 4;  Length 503;
Best Local Similarity  24.7%;  Pred. No. 6.6e-16;
Matches 136;  Conservative 91;  Mismatches 204;  Indels 120;  Gaps 24;

Qy  303  EP-ELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIV  361
                                                                       
Db   12  EPGAVERRDYQVGLAEQAIR-ENCIVLPTGLGKTAVALQVISHYLDEGRGALFLAPTRV  70

Qy  362  LVNKVLLVEQLFRKBFQPPLKKWYRVIGL-SGDTQL--KISFPEVVKSC--DIIISTAQIL  417
                                                                        
Db   71  LVN------QHRQFLG---RALTISDITLVTGEDTVPRRKKAWGSVICATPEIT  116

Qy  418  ENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKP  477
                                                                       
Db  117  RNDIAR------GMVPLEQFGLVVFDEAHRAVGDYAYSAIA-------RAVGENS-  158

Qy  478  VIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCK  537
                                                                       
Db  159  ----RMIGMTATL---PSEREKADE----IMGTLLSKSIAQRTEDDPVKPYVQETET  205

Qy  538  KFAIADATREDPFFKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRK  596
                                                                        
Db  206  EWIKVELPEMKEIQKLLK-MALDERYAALKRCGYDLGSNRSLSALLRL-RMVVLSGNRR  263

Qy  597  ERVCAEHLRKYNEALQINDTIRMIDAY--THLETFYNEBKDKKFAVIEDDSDEGGDDEYC  654
                                                                       
Db  264  A------AKPLFTAIRITYALNIFEAHGVTPFLKFCERIVKKGAGVAE---------  306
```

```
Db   342  --RGARGKALVFTSYRDSVDLI------HSRLKAAGINSGILIGKAGE---KGLKQRKQ  389
Qy   772  KEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR-ARADESTY  830
Db   390  VETVAKFRDGGYDVLVSTRVGEEGLDISEVNLVIFYDNVPSSIRYVQRRGRTGRKDAGRL  449

Qy   831  VLVAHSGS  838
Db   450  IVLMAKGT  457
```

RESULT 3
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Väisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

```
Query Match      4.1%;  Score 217.5;  DB 4;  Length 2662;
Best Local Similarity  18.6%;  Pred. No. 7.6e-09;
Matches 209;  Conservative 185;  Mismatches 421;  Indels 307;  Gaps 47;

Qy     9  ENFRYLISCFRARVKMYIQVZPVLDYLTFLPABVKEQIQRTVATSGMQAVELLLSTLEK    68
Db  1039  EQQRKIFSLIQEKNELQQMLSSVIAEKEQLKTDLKENIEMTIB---NQEELRLLGDELKK  1095

Qy    69  GVWHLGWTREFV-----EALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNLLQP   123
Db  1096  -----QOEIVAQEKNHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQLLNVQEE    1148

Qy   124  TLVDKLLVRDV--LDKCMEERLLTIE--DRNRIAAAEN-NGNESGVRELLK-RIVQKE-    175
Db  1149  MSEMQKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKBRKVLKELQ  1208

Qy   176  --------NWFSAFLNVLRQTG-----NNELVQELTGSDCSESNAEIENL            212
Db  1209  KSFETBRDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRS-VSEKTAQIINT  1267

Qy   213  SQVDGPQV----------EEQLLSTTVQPNLEKEVWGMENNSSESSFADSVVSESDTS    261
Db  1268  QDLEKSHTKLQEEIPVLHEEQELL-----PNVKKVSETQETMNELELLTEQSTTKDS-TT  1321

Qy   262  LAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARASPPELQLRPYQMEVAQPALE   321
Db  1322  LARIEM---ERLRLNEKF--------QESQEEIKSLTKERDNLKTIKEALEVKHDQL-   1367

Qy   322  GKNIIICLPTGSGKTRVAVYIAK---------DHLDKKKASEPGKVILVNKVLLVEQL   372
Db  1368  -----------KEHIRETLAKIQESQSKQEQSLNMKEKDNETTKI-----VSEMEQ    1407

Qy   373  FRKEFQPFLKKWYRVIGLSGDTQLKISFPE---VVKSCDIIISTAQILENSLLNLENG-   427
Db  1408  FKPKDSALLRIEIEMLGLS--KRLQESHDEMKSVAKEKDDLQRLQESEDQLKENIK    1465
```

```
Qy   655  DGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTR  714
Db   307  -----------------LFEEDR-----NFTGAMARAKAQAAGMEHPKIPKL--------  337
Qy   715  TEESARG-----IIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIG-AGHSSEFKPMTQ  768
Db   338  -EEAVRGAKGKALVFTSYRDSVDLI------HSKLQAAGINSGILIGKAGE---KGLKQ   386
Qy   769  NEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR-ARADE  827
Db   387  KKQVETVAKFRDGGYDVLVSTRVGEEGLDISEVNLVIFYDNVPSSIRYVQRRGRTGRKDA  446

Qy   828  STYVLVAHSGS  838
Db   447  GKLVVLMAKGT  457
```

RESULT 2
US-09-408-020-34
; Sequence 34, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCORP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SEQ ID NO 34
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-34

```
Query Match      5.3%;  Score 284;  DB 4;  Length 502;
Best Local Similarity  24.3%;  Pred. No. 9.8e-16;
Matches 133;  Conservative 94;  Mismatches 207;  Indels 114;  Gaps 23;

Qy   303  EP-ELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIV   361
Db    12  EPGAVERRDYQVGLABQAIR-ENCIVLPTGLGKTAVALQVIAHYLDEGRGALFLAPTRV    70

Qy   362  LVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQL--KISFPEVVKSC--DIIISTAQIL  417
Db    71  LVN--------QHRQFLG--RALTISDITLVTGEDTIPRRKKAWGGSVICATPEIA     116

Qy   418  ENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKP  477
Db   117  RNDI------ERGLVPLEQFGLVIFDEAHRAVGDYAYSSIAR------------A     153

Qy   478  VIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCK  537
Db   154  VGDNSRMVGMTATL----PSEREKADE----IMGTLLSRSIAQRTEDDPDVKPYVQETAT   205

Qy   538  KFAIADATREDPFKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRK  596
Db   206  EWIKVDLPPEMKEIQRLLKALD-ERYSSLKRCGYDLGSNRSLSALLRL-RMVVLGGNRR   263

Qy   597  ERVCAEBHLRKYINEALQINDTIRMIDAY--THLETFYNEEKDKKFAVIEDDSDEGGDDEYC  654
Db   264  A------AKPLFTAIRITYALNIFEAHGVTPLKFCERTSKKKGVGVAE---------    306

Qy   655  DGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTR  714
Db   307  -----------------LFFQDR-----NFTGAIARAKAQAAGMEHPKIPKLEDAV--   341

Qy   715  TEESARG--IIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIG-AGHSSEFKPMTQNEQ  771
Db   ...
```

```
Qy    57 QAVELLLSTLEKGVWHLGWTREFVEALRTGSPLAARYMNPEL-TDLPSPSFENAHDEYL   115
            :: :::   :  : :   :  : :.:       :   :      :  ::
Db  1194 KPLHLLIGKLQKAV------SEECSYFLQTLCSVLGEYYTPALKCEVNAEDKENSGD---  1244

Qy   116 QLLNLLQPTLVD-KLLVRD------VLDKCMEE--ELLTIEDR-NRIAAAENNGN--ES   162
            ::: ::  :     : :           :  :    ::  ::.: :
Db  1245 YISENEDEPELQDYRYEVQDFQENMHTLLNKVTEYNKLLVLQTRLSKIWGQQTDGMKLEF  1304

Qy   163 GVRELLKRIVQKENWFSAFLNVLRQTGNNE------------LVQELTGSDCSESNAEIEN   211
            :  :::            :   :    :|||:           :
Db  1305 GEENLPKEETE--------FLSIHSQMTNLEDIDVNHKSKLSSLQDLEKTKLEEQVQELES  1357

Qy   212 LSQVDGPQVEEQLLSTTVQPNLKEKEVWGMENN---SSESSF---ADSSVVSESDTSLA   263
            :          ::::      :   :   :.:      ::::: ||| :::
Db  1358 LIS----SLQQQLKET-EQNYEAEIHCLQKRLQAVSESTVPPSLPVDSVVITESDAQRT   1411

Qy   264 EGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARASPEPELQLRPYQMEVAQPALEGK   323
                       :  :      :  :       :        :  :::::.      : ::
Db  1412 MYPGSCVKKNIDGTIEFSSGEFGV---KEETNIV----------KLLEKQYQEQLEE--EVA   1457

Qy   324 NIIICLP---------------------TGSGKTRVAYIAKDHLDKKKASEPGKVIVLVNK   365
            : ::::                       ::::::::   ::  : :  ::: : :::.:
Db  1458 KVIVSMSIAFAQQTELSRISGGKENTASSKQAHAVCQQEQHYFNEMKLSQDQ--IGFQT   1514

Qy   366 VLLVEQLFRKEFQPFLKKWYRVIGLSG-DTQLKISFPEVV---KSC               407
            : :::::|:::|:|::::: ||||:::   :: |::::
Db  1515 FETVDVKFKEEFKPLSKE---LGEHGKEILLSNSDPHDIPESKDCVLTISEEMPSKDKT   1570

Qy   408 ------------DIILSTAQILENSLLNLNENGEDAGVQL---------SD            436
                         :::::::::::::::
Db  1571 FIVRQSIHDEISVSSMDASRQLMNEEQLEDMRQELVRQYQEHQQATQRSSIDNENLVSE   1630

Qy   437 FSLIIIDE---------------------CHHTNKEAVYNN                     456
            :::::::                       ||||:|||:::
Db  1631 RERVLLELELEALKQLSLAGREKLCCELRNSSTQTQNGNENQGEVEEQTFKEKELDRKPED  1690

Qy   457 -----IMRHYLMQKLKNNRLKKENKPVIPLPQILGLTASPGVGG---               495
                  |||||||||||||||||
Db  1691 VPPEILSNERYALQK-ANNRLLKILLEVVKTTAAVEETIGRHVLGILDRSSKSQSSASLI  1749

Qy   496 ------ATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKK-------F     539
                  |||||||||||||||||||||||||||||
Db  1750 WRSEAEASVKSCVHEEHTRVTDESIPSYSGSDMPRNDINWSKVTEEGTELSQRLVRSGF   1809

Qy   540 AIADATREDPFKEKL-LEIMTRIQTYCQ--MSPMSDFGTQ------PYEQWAIQMEKKAAKK   592
            |||||||||||||||  :::|||||||   ||:|||||.
Db  1810 A---GTEIDPENEELMLNISSRLQAAVEKLLEAISETSSQLEHAKVTQTELMRESFRQKQ   1866

Qy   593 GNRKERVCAEHLRK--YNE----------ALQINDTIRMIDAYTHLETFYNEBKDKKFAVI-   641
             :..: |      :   ::           : :   .  :  :
Db  1867 EATESLKCQEELRERLHEESRAREQLAVELSKAEGVIDGYADRKTLFERQIQEKTDIIDR  1926

Qy   642 -----------------------------------EDDSDEGGDDE----------     652
                                               |:::|:|||:
Db  1927 LEQELLCASNRLQELEAEQQIQEERELLSRQKEAMKARAGPVEQQLLQETEKLMKEKLE  1986

Qy   653 -YCDGDEDEDDLKKPLKLDETD--------RFLMTLFFENNKMLKRLAENPEYENEKLTKLR   705
             ::         :  :|:|:|         |||:|||||||||||||:|||:||||||||:
Db  1987 VCCQAEKVRDDLQKQVKALBIDVEEQVSRFI-ELEQEKNTELMDLRQQNQALEKQLEKMR   2045

Qy   706 NTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKP   765
            ::::|| :||||
Db  2046 KFLDEQAIDREHERD--------------------VFQQFIQKLEQQLKV            2081

Qy   766 MTQNEQKEVIS------KFRTGKINLLLIATTVAEEGL--DIKECNIVIRYGLVTNEIAMVQ   818
            ::|:|| |:       |||||||||:||||:|||:::  |||||||||||||:||:
Db  2082 ISEHQTREVEQLANHLKEKTDKCSELL--LSKEQLQRDIQERN-------EEIEKLE      2129

Qy   819 ARGRARADESTYVLVAHSGS--------------------------------GVIE       842
                       :|||:||:||                                |:|:
Db  2130 FRVR----ELEQALLVSADTFQKVEDRKHFGAVEAKPELSLEVQLQAERDAIDRKEKEIT   2185

Qy   843 --HETVNDFREKOMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKR----NIAKH   895
```

```
Qy   428 EDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIIGL   487
            :         ::: :: ::::|::::: ::    :
Db  1466 EIVAKHLETEEELKVAHCCLKEQEETIN----------ELRVVNLSEKETE----------  1505

Qy   488 TASPGVGGATKQAKA------EEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKF    539
            :            :        :::  : |: :     :::|:::   :|::
Db  1506 ----ISTIQKQLEAINDKLQNKIQEIYEKEEQLNIKQISEVQENVNELK-QFKEHRK-    1557

Qy   540 AIADATREDPFKEKLLEIMTRIQTYCQ---MSPMSDFGTQPYEQWAIQMEKKAAKKGNR   595
            | ||:|         :|:  :::|:::   ::    :   :    :  :::
Db  1558 --AKDSALQSIESKMLELTNRLQESQEEIQIMKEKEMKRVQEALQIERDQLKENTKEI   1615

Qy   596 ----------KERVC-ABHLRKYNEALQIN------DTIRMID-AYT              624
                      :::    |:||:: :  : :       :::::.
Db  1616 VAKMKESQEKYQFLKMTAVNETQEKMCEIEHLKEQFETQKLNLENIETENIRLTQILHE  1675

Qy   625 HLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLD----------   670
            :|:    :  |: :|:|  :|:|||:|| ||||:|:::|::|:|::
Db  1676 NLEEMRSVTKER----------DDLRSVEETIKVERDQLKENLRETITRDLEKQEELKIVHMHL  1729

Qy   671 ----ETDRFLMTLFFENNKMLKRLAENPEYENEKL------------TKLRNTIMEQYRTE   716
                |:|: | ::  ::: ::::: :    :   :             |||:|:|||||:|
Db  1730 KEHQEETDKLRGIVSEKTNEISNMQKDLEHSNDALKAQDLKIQEELRIAHMHLKEQQETI   1789

Qy   717 ESARGIIFTKTRQSAYALSQWITEN-----EKPAEVGVKAHHLIGAGHS-------SEFK   764
            | :   ::  | :   : |  : :       ||  || :  ::: : |:       ::::
Db  1790 DKLRGIVSEKTDKLSNMQKDLENSNAKLQEKIQELKANEHQLITLKKDVNETQKKVSEME   1849

Qy   765 PMTQ--NEQKEVISKFRTGKINLLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR   822
            |::|   |::  : :  : ::   ::: :     : ::|   ::   :
Db  1850 QLKKQIKDQSLTLSKLEIENLNL--AQELHENLEEMK---------------SVMKERDN  1892

Qy   823 ARADESTYVLVAHSGSGVI------------------------------EH-ETVNDFREKMMYKAIH   859
            : |:|   : :   :: :                              |:  |||||:|:::::|
Db  1893 LRRVEETIKLERDQLKESLQETKARDLEIQQELKTARMLSKEHKETVDKLREKISERTIQ   1952

Qy   860 CVQNMK-----PEEYAHKILELQMQSIM----------EKKMKTKRNIAKHYKNNPSLI    903
                      || :   :::|::|:::             :: ::: :
Db  1953 ISDIQKDLDKSKDELQKKIQELQKKELQLLRVKEDVNMSHKKINEMEQLKKQFEPN----  2008

Qy   904 TFLCKNCSVLACSGEDIHVIEKMHHVAMTPEFKELYIVRENK                    945
             :|||          ||  ::|:   : :       : ::: |
Db  2009 -YLCK------CEMDNFQLTKKLHE----SLEEIRIVAKER                     2038
```

RESULT 4
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match   3.8%; Score 202; DB 4; Length 3878;
Best Local Similarity   18.8%; Pred. No. 3e-07;
Matches 238; Conservative 179; Mismatches 424; Indels 428; Gaps 55;

```
Qy     5 YSTD-------ENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNM    56
         :::.        |:||:::|||||:|::||:|:||||:||  :|:||||::  :
Db  1136 YSTHVDQVREYMENEKDKALCSLKEELIFAQEEKIKELQKIHQLEL--QTMKTQETGDEG  1193
```

```
Db   1355 KTQELESHQSECLHCIQVAEAEVKEKTELLQTLS-SDVSELLKDKTHLQEKLQSLEKDSQ 1413
Qy    224 LLSTT-------VQPNLEKEVVWGMENNS-----SESSFADSSVVSESDTSLAEGS---- 266
Db   1414 ALSLTKCELENQIAQLNKEKELLVKESESLQARLSESDYEKLNVSKALEAALVERGEFAL 1473
Qy    267 -VSCLDESLGHNSNMGSDSGTMGSDSDE-------ENVAARASPEPELQLRPYQMEVAQP 318
Db   1474 RLSSTQEEV-HQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSLKDKVENLERELQ 1532
Qy    319 ALEGKNIIICLPTGSGKTRVAVYIAK-DHLDKKKKASEPGKVIVLVNKVLLVEQLFRKEF 377
Db   1533 MSBENQELVILDAENSKAEVETLKTQIEEMARSLKVFELDIVTLRSEKENLTKQIQEKQG 1592
Qy    378 QPFLKKWYRVIG-----LSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLEN----- 426
Db   1593 Q--LSELDKLLSSFKSLLEEKQAEIQIKEESK----TAVEMLQNQLKELNEAVAALC 1644
Qy    427 GEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQI- 484
Db   1645 GDQEIMKATEQSLDPPIEEEHQLRNSI-------EKLRARLEADEKKQLCVLQQLKE 1694
Qy    485 --------LGLTASPGVGGATKQAKAEEHILKL---------CANLD 514
Db   1695 SEHHADLLKGRVENLERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQSLRGLELD 1754
Qy    515 AFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFG 574
Db   1755 VVTIRSEKENLTNELQKEQERISELEIINSSFENILQEK----EQEKVQMKEKSSTA 1807
Qy    575 TQPYEQWAIQMEKKAAKKGNRKERVCA--EHLRKYNEALQIN---DTIRMIDA----YTHL 626
Db   1808 MEMLQTQLKELNERVAALHNDQEACKAKEQNLSSQVECLELEKAQLLQGLDEAKNYIVL 1867
Qy    627 ETFYN---------EEKDKKFPAVIEDDSDEGGDDEYCDGDEDDLKKPLKLDETD 673
Db   1868 QSSVNGLIQEVEDGKQKLEKKDEEISRLKNQI---------QDEQLVSKLSQVEGE 1915
Qy    674 RFLMTLFFENNKMLKRLAENPEY------ENEKLTKLRNTIMEQYTRTEESARGIIFTK 726
Db   1916 H---QLWKEQNLELRNLTVELEQKIQVLQSKNASLQDTLEVLQSSYKNLENE--LELTK 1969
Qy    727 TRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQEVISKFRTGKINLL 786
Db   1970 MDKMSFV---------EKVNKMTAKETELQREMHEMAQKTAELQEELSGEKNRLAGELQLL 2021
Qy    787 ---------IATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYV 831
Db   2022 LEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEIABYQLRLHEAEKKHQAL 2081
Qy    832 LVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPE------BYAHKILELQMQSIME 883
Db   2082 LLDTNKQYEVEIQT--YREKLTSKE-ECLSSQKLEIDLLKSSKELNNSLKATTQ-ILE 2136
Qy    884 KKMKTK---------RNIAKHYKQNPSLITFLCKNCSVLACSGEDIHVIEK------ 925
Db   2137 ELKKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLE--EEKEILQKELSQLQAAQEK 2193
Qy    926 -------MEHVNMTPEFKELYIVRENKA-----LQKKCADYQINGBIICKCCGQAWGTM 971
Db   2194 QKTGTVMDTKVDELTTEIKELKETLEEKTKEADEYLDKYCS-LLISHEKLEKAKEMLETQ 2252
Qy    972 MVH 974
Db   2253 VAH 2255
```

RESULT 6
US-09-107-532A-3677
; Sequence 3677, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush

```
Db   2186 NLEEQLEQFREEL----------ENKNEEVQQLHMQLEIQKKESTTRLQEBLEQENKL 2232
Qy    896 YKNNPSLITFLCKNCSVLACSGEDIHVI-----EKMHHVN----MTPEFKELYIV 941
Db   2233 FKDDMEKLGLAIKESD--AMSTQDQHVLFGKFAQIIQEKEVEIDQLNEQVTKLQQQLKIT 2290
Qy    942 RENKALQKK 950
Db   2291 TDNKVIEEK 2299
```

RESULT 5
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match   3.7%; Score 194.5; DB 1; Length 2482;
Best Local Similarity 19.2%; Pred. No. 6.5e-07;
Matches 219; Conservative 186; Mismatches 473; Indels 265; Gaps 46;

Qy   21 RVKMYIQVEPVL----------DYLTFLPAEVKEQIQRTVATSG-----NMQAVELLLS 64
Db   1189 KIEACIELEKIVGELKKENSDLSEKLEYFSCDHQELLQRVETSEGLNSDLEMHADKSSRE 1248
Qy   65 TLEKGVWHL--GWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNLLQ 122
Db   1249 DIGDNVAKVNDSWKRFLD----------VENELSRIRSEKASIEH-EALYLEADLE 1294
Qy   123 PTLVDKL-LVRDVLDK-----CMEEELLTI-EDRNRIAAAENNGNE--SGVRELLKRIVQ 173
Db   1295 VVQTEKLJCLEKDNENKQKVIVCLEEELSVVTSERNQLRGELDTMSKKTTALDQLSEKMKE 1354
Qy   174 K-----ENWFSAFLNVLRQT---GNNELVQELTGSDCSESSNABIENLSQ-VDGPQVEEQ 223
```

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
    ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
   ADDRESSEE: GENOME THERAPEUTICS CORPORATION
   STREET: 100 Beaver Street
   CITY: Waltham
   STATE: Massachusetts
   COUNTRY: USA
   ZIP: 02354
COMPUTER READABLE FORM:
   MEDIUM TYPE: CD/ROM ISO9660
   COMPUTER: PC
   OPERATING SYSTEM: <Unknown>
   SOFTWARE: ASCII
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/107,532A
   FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 60/085,598
   FILING DATE: 14 May 1998
   APPLICATION NUMBER: 60/051571
   FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
   NAME: Ariniello, Pamela Deneke
   REGISTRATION NUMBER: 40,489
   REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (781)893-5007
   TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3677:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 661 amino acids
   TYPE: amino acid
   TOPOLOGY: linear
   MOLECULE TYPE: protein
   HYPOTHETICAL: YES
   ORIGINAL SOURCE:
   ORGANISM: Enterococcus faecium
   FEATURE:
   NAME/KEY: misc feature
   LOCATION: (B) LOCATION 1...661
   SEQUENCE DESCRIPTION: SEQ ID NO: 3677:
US-09-107-532A-3677

```
Query Match          3.6%;    Score 193.5;  DB 4;    Length 661;
Best Local Similarity  18.8%;  Pred. No. 9.4e-08;
Matches 147; Conservative 130; Mismatches 240; Indels 267; Gaps 36;

Qy  299  RASPEPELQLRPYQMEVAQP---------ALEGKNIIICL-PTGSGKTRVAVYIAKDHLD 348
              : :::  ::::     :   ||||||  ::|||||
Db    2  RASRHFEL-VSKYQPAGDQPEAINQLVDGVVGGKKAQILLGATGTGKT---YTISNLIE  56

Qy  349  KKKKASEPGKVIVLVNKLLVEQLFR--KEFQPFLKKWYRVIGLSGDTQLKISF-----P 401
              :: :: |||        ::::       ||||
Db   57  KVNKPT--LIIAHNKTLAGQLYGEFKEFFP---------NNAVEYFVSYDYYQP 100

Qy  402  EV-VKSCDIIISTAQILENSLLNLENGEDAG-VQLSDFSLIIIDEC------HHTNKEA 452
             :|:  :::  :: :             :
Db  101  EAYVPSSDTYIEKDSSVNDEIDKLRHSATSSLLERNDVIVIASVSCIFGLGSPFEYQKQV 160

Qy  453  V-------YNNIMRHYLMQKLKNN------RLKKENKPVIPLP--------------- 482
          :|:                       :
Db  161  VSIRQGAELDRNQLIRDLVSIQFERNDIDFQRGRFRVRGDVVEIFPASRDERALRVEFFG 220

Qy  483  --------QILGLTASPGVGGATKQAKAREHILKLCANLDAFTIKTVKENLDQL 528
                          :|:    ||::
Db  221  DEVERIREVNALTGEVLGETEHVAIFPATHPVTNDEHMEHAVANIKA---ELEQRLTVL 276

Qy  529  KNQIQEPCKKFAIADATREDFFKEKLLEIMTRIQ-TYCQMSPMSDFG-TQPYEQWAIQME 586
            |:                      ::::  :    :: ::::
Db  277  RNE-------------NKLLEAQRLEQRTNYDIEMMLEMGYITSGIENYSRHMD 316
```

```
Qy  587  KKAAKKGNRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSD 646
               :  :|:                           :|: :|:  |:  ::|:|:
Db  317  GR-KEGE----------------------------PPYTLLD-FFPED---FLIVADESH 343

Qy  647  ------EGDDEYCDGDEDDL--------------KKPLKLDETDRFLMTLFFENNKML 687
                                                ::|| :|:  :::
Db  344  VTMPQIRG---MYNGDRARKQMLVDYGFRLPSALDNRPLRLEFFEKHV-------NQII 392

Qy  688  KRLAENPEYENEK--------------------LTKLRNTIMEQYTRTEESARGII 723
            ||||                              : ::  :::: ::|:
Db  393  YVSATPGPYEHEQTDTVIQQIIRPTGLLDPVIEVRPIMGQIDDLVGEINERVEKDQRVFV 452

Qy  724  FTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKI 783
            :|:  |:|:                       ::  :||:|:
Db  453  TTLTKKMABPLTDY------FKELGIKVKYL-----HSDIKTL--ERTEIIRDLRLGEF 498

Qy  784  NLLIATTVABEGLDIKECNIVI-----RYGLVTNEIAMVQARGRARADESTYVLVAHSGS 838
            ::|:   :||:|||               ::: ::|:
Db  499  DVLVGINLLREGLDVPEVSLVAILDADKEGFLRSERSLVQTIGRAARNEBGKVIM----- 553

Qy  839  GVIEHETVNDFREKMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKN 898
                                               ::|::
Db  554  --------------------------YADKVTD-SMRLAMDETSR-RRTIQQKYNE 581

Qy  899  N------PSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQ--KKC 951
            :                          :|:|   : :
Db  582  EHGIVPKTTIKEIRDLISITKESEDDTKEAVQVSYEBMTKEEKDTLLMKLEKEMKDAAKA 641

Qy  952  ADYQ 955
          :|:
Db  642  LDFE 645
```

RESULT 7
US-08-353-700-1
  Sequence 1, Application US/08353700
  Patent No. 5599919
  GENERAL INFORMATION:
   APPLICANT: YEN, TIMOTHY J.
   APPLICANT: RATTNER, JEROME B.
   TITLE OF INVENTION: NUCLEIC ACID ENCODING A
   TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
   TITLE OF INVENTION: AND METHODS OF USE
   NUMBER OF SEQUENCES: 4
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
   STREET: 1601 MARKET STREET, SUITE 720
   CITY: PHILADELPHIA
   STATE: PA
   COUNTRY: USA
   ZIP: 19103-2307
   COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/353,700
   FILING DATE: 09-DEC-1994
   CLASSIFICATION: 435
   ATTORNEY/AGENT INFORMATION:
   NAME: REED, JANET E.
   REGISTRATION NUMBER: 36,252
   TELECOMMUNICATION INFORMATION:
   TELEPHONE: (215) 563-4100
   TELEFAX: (215) 563-4044
   INFORMATION FOR SEQ ID NO: 1:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 3248 amino acids
   TYPE: amino acid
   STRANDEDNESS: single
   TOPOLOGY: linear
   MOLECULE TYPE: protein

```
;  HYPOTHETICAL: NO
;  ANTI-SENSE: NO
;  ORIGINAL SOURCE:
;    ORGANISM: HUMAN
US-08-353-700-1

Query Match         3.6%;  Score 191.5;  DB 1;  Length 3248;
Best Local Similarity  19.2%;  Pred. No. 1.8e-06;
Matches 221;  Conservative 184;  Mismatches 465;  Indels 281;  Gaps 47;

Qy    21 RVKMYIQVEPVL--------DYLTFLPAEVKEQIQRTVATSG-----NMQAVELLS 64
              :  :::        |||||||||       ::||:
Db  1917 KIEACIELEKIVGELKKENSDLSEKLEYFSCDHQELLQRVETSEGLNSDLEMHADKSSRE 1976

Qy    65 TLEKGVWHL--GWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNLLQ 122
             :  : :     |  |::: |:
Db  1977 DIGDNVAKVNDSWKERFLD----------------VENELSRIRSEKASIEH-EALYLEADLE 2022

Qy   123 PTLVDKL-LVRDVLDK-------CMEEELLTI-EDRNRIAAAENNGNE--SGVRELLKRIVQ 173
             :|||:|  :|::||   :: |||       ::
Db  2023 VVQTEKLCLEKDNENKQKVIVCLEEELSVVTSERNQLRGELDTMSKKTTALDQLSEKMKE 2082

Qy   174 K----ENWFSAFLNVLRQT--GNNELVQELTGSDCSESNAEIENLSQ-VDGPQVEEQ 223
              |||:|||    :: |:|
Db  2083 KTQELESHQSECLHCIQVAEAEVKEKTELLQTLS-SDVSELLKDKTHLQEKLQSLEKDSQ 2141

Qy   224 LLSTT------VQPNLEKEVWGMENNS-----SESSFADSSVVSESDTSLAEGS------ 266
             ||:|         :|:: ::
Db  2142 ALSLTKCELENQIAQINKEKELLVKESESLQARLSESDYEKLNVSKALEARALVEKGEFAL 2201

Qy   267 -VSCLDESLGHNSNMGSDSGTMGSDSDB-------ENVAARASPEPELQLRPYQMEVAQP 318
              :  |:: :                    |::
Db  2202 RLSSTQEEV-HQLRRGIEKLRVRIFEADEKKQLHIAEKLKPERERENDSLKDKVENLERELQ 2260

Qy   319 ALEGKNIIICLPTGSGKTRVAVYIAK-DHLDKKKKRASEPGKVIVLVNKVLLIVEQLFRKEF 377
             ::|  : :: |
Db  2261 MSEENQELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQG 2320

Qy   378 QPFLKKWYRVIG-----LSGDTQLKISFPEVVKSCDIIISTAQILENSLINLEN----- 426
             |  :|:|    |||      :::|
Db  2321 Q--LSELDKLLSSFKSLLEEKEQAEIQIKEESK-------TAVEMLQNQLKEINEAVAALC 2372

Qy   427 GEDAGVQVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQI-- 484
             |::      |          ::|||     :|
Db  2373 GDQEIMKATEQSLDPPIEEEHQLRNSI------------EKLRARLEADEKKQLCVLQQLKE 2422

Qy   485 ------LGITASPGVGGATKQAKAEEHILKL---------CANLD 514
              |  |::|  ||  ||          |:
Db  2423 SEHHADLLKGRVENLERELEIARTNQEHAALEAENSKGEVETLKAKIEGMTQSLRGLELD 2482

Qy   515 AFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFG 574
             |: |:|:|:|:|  ||            :    : ||
Db  2483 VVTIRSEKENJTNELQKEQERISELEIINSSFENILQEK------------- 2521

Qy   575 TQPYEQWAIQMEKKAA---------KKGNRKERVCAEH-------LRKYNEALQIN-- 614
             ::   : ::|||||            :    ::                 :|:||:|:
Db  2522 ---EQEKVQMEKKSSTAMEMLQTQLKELN-ERVAALHNDQEACKAKEQNLSSQVECLE 2575

Qy   615 -DTIRMIDAYTHLETFYNEEKDKKFAVIEDDSD-----EGGDDEYC---DGDEDEDDIKK 665
                  :    |:                     : ||      :      :|:|:|:|:|
Db  2576 LEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGKQKLEKKDEEISRLKNQIQDQEQLVS 2635

Qy   666 PLKLDETDRFLMTLFFENNKMLKRLAENPEY-------ENEKLTKLRNTIMEQYTRTEES 718
             |:           | :::     | ::         : |::|:|:|:
Db  2636 KLSQVEGEH--QLWKEQNLELRNLITVELEQKIQVLQSKNASIQDTLFVLQSSYKNLENE 2692

Qy   719 ARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKF 778
             :  ::    :     :
Db  2693 --LELTKMDKMSFV-------EKVNKMTAKETELQREMHEMAQKTAELQEELSGEKNR 2741

Qy   779 RTGKINLL-------IATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRA 823
                : :||       :  :|||:
Db  2742 LAGELQLLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEIAEYQLRLHE 2801
```

```
Qy   824 RADESTYVLVAHSGSGVIEHETVNDFREKMYKAIHCVQNMKPE------EYAHKILE 875
             :  :  :|  :          :   :||:|       |:
Db  2802 AEKKHQALLLDTNKQYEVEIQT---YREKLTSKE-ECLSSQKLEIDLLKSSKEELNNSLK 2857

Qy   876 LQMQSIMEKKMKTK---RNIAKYHYKNNPSLITFLCKNCSVLACSGEDIHVIEK--- 925
             :: :|:|:||:|:|   |                    ::|::
Db  2858 ATTQ-ILEELKKTKMDNLKVVNQLKKENERAQGKMKLLIKSCKQLE--EEKBIILQKELS 2913

Qy   926 --------MHHVNMTPEFKELYIVRENKA-----LQKKCADYQINGEIICK 963
                      : :|||||     ::                 :||
Db  2914 QLQAAQEKQKTGTVMDTKVDELTTEIKELKELEEKTKEADEYLDKYCS-LLISHEKLEK 2972

Qy   964 CGQAWGTMMVH 974
             ::  || | |
Db  2973 AKEMLETQVAH 2983
```

```
RESULT 8
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
;    APPLICANT: Yen, Timothy J.
;    APPLICANT: Rattner, Jerome B.
;    TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
;    TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
;    NUMBER OF SEQUENCES: 4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Dann, Dorfman, Herrell and Skillman
;      STREET: 1601 Market Street   Suite 720
;      CITY: Philadelphia
;      STATE: PA
;      COUNTRY: USA
;      ZIP: 19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: PCT/US95/16216
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/353,700
;      FILING DATE: 09-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Reed, Janet E.
;      REGISTRATION NUMBER: 36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (215) 563-4100
;      TELEFAX: (215) 563-4044
;    INFORMATION FOR SEQ ID NO: 1:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: 3248 amino acids
;        TYPE: amino acid
;        STRANDEDNESS: not relevant
;        TOPOLOGY: not relevant
;      MOLECULE TYPE: protein
;      HYPOTHETICAL: NO
;      ANTI-SENSE: NO
PCT-US95-16216-1

Query Match         3.6%;  Score 191.5;  DB 5;  Length 3248;
Best Local Similarity  19.2%;  Pred. No. 1.8e-06;
Matches 221;  Conservative 184;  Mismatches 465;  Indels 281;  Gaps 47;

Qy    21 RVKMYIQVEPVL--------DYLTFLPAEVKEQIQRTVATSG-----NMQAVELLS 64
              :  :::        |||||||||       ::||:
Db  1917 KIEACIELEKIVGELKKENSDLSEKLEYFSCDHQELLQRVETSEGLNSDLEMHADKSSRE 1976

Qy    65 TLEKGVWHL--GWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNLLQ 122
```

```
Db  1977 DIGDNVAKVNDSWKERFLD----------VENELSRIRSEKASIEH-EALYLEADLE 2022

Qy   123 PTLVDKL-LVRDVLDK----CMEEELLTI-EDRNRIAAAENGNE--SGVRELLKRIVQ 173

Db  2023 VVQTEKLCLEKDNENKQKVIVCLEEELSVVTSERNQLRGELDTMSKKTTALDQLSERMKE 2082

Qy   174 K----ENWFSAFLNVLRQT----GNNELVQELTGSDCSESNAEIENLSQ-VDGPQVVEQ 223

Db  2083 KTQELESHQSECLHCIQVAEAEVKEKTELLQTLS-SDVSELLKDKTHLQEKLQSLEKDSQ 2141

Qy   224 LLSTT----VQPNLEKEVWGMENNS-----SESSFADSSVVSESDTSLAEGS---- 266

Db  2142 ALSLTKCELENQIAQLNKEKELLVKESESLQARLSESDYEKLNVSKALEAALVEKGEFAL 2201

Qy   267 -VSCLDESLGHNSNMGSDSGTMGSDSDE----ENVAARASPEPELQLRPYQMEVAQP 318

Db  2202 RLSSTQEEV-HQLRRGIEKLRVRIEADEKKQLHIAEKLKERERENDSLKDKVENLERELQ 2260

Qy   319 ALEGKNIIICLPTGSGKTRVAVYIAK-DHLDKKKKASEPGKVIVLVNKVLLVEQLFRKEF 377

Db  2261 MSEENQELVILDAENSKAEVETLKTQIEMARSLKIFELDLVTIRSEKENLTKQIQEKQG 2320

Qy   378 QPFLKKWVRIG----LSGDTQLKISPPEVVKSCDIIISTAQILENSLLNLEN------ 426

Db  2321 Q--LSELDKLLSSFKSLLEEKEQAEIQIKEESK-----TAVEMLQNQKELNEAVAALC 2372

Qy   427 GEDAGVQLSDFSLIIIDECHHTNKEAVTNNIMRHYLMQKLKNNRLKKENKPVIPLPQI- 484

Db  2373 GDQEIMKATEQSLDPPIEEHQLRNSI-----EKLRARLEADEKKQLCVLQQLKE 2422

Qy   485 ------------------LGLTASPGVGGATKQAKAEEHILKL---------CANLD 514

Db  2423 SEHHADLLKGRVENLEREILEIARTNQEHAALEAENSKGEVETLKAKIEGMTQSLRGLEID 2482

Qy   515 AFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFFKSKLLEIMTRIQTYCQMSPMSDFG 574

Db  2483 VVTIRSEKENLTNELQKEQERISLEIINSSFENILQEK--------- 2521

Qy   575 TQPYEQWAIQMEKKAA---------LRKYNEALQIN--- 614

Db  2522 ---EQEKVQMKEKSSTAMEMLQTQLKELN-ERVAALHNDQEACKAKEQNLSSQVECLE 2575

Qy   615 -DTIRMIDAYTHLETFYNEEKDKKFAVIEDDSD-----EGGDDEYC--DGDEDEDDLKK 665

Db  2576 LEKAQLLQCLDEAKNNYIVLQSSVKGLIQEVEDGKQKLEKKDEEISRLKNQIQDQEQLVS 2635

Qy   666 PLKLDETDRFLMTLFFENNKMLKRLAAENPEY--------ENEKLTKLRNTIMEQYTRTEES 718

Db  2636 ---KLSQVEGEH--QLWKEQNLELRNLTVELEQKIQVLQSKNASLQDTLEVLQSSYKNLENE 2692

Qy   719 ARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSHSSEFKPMTQNEQKEVISKF 778

Db  2693 --LELTKMDKMSFV---------EKVNKMTAKETELQREMHEMAQKTABLQEELSGEKNR 2741

Qy   779 RTGKINLL----------IATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRA 823

Db  2742 LAGELQLLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEIAEYQLRLHE 2801

Qy   824 RADESTYVLVAHSGSGVIEHETVNDFREKXMYKAIHCVQNMKPE----------EYAHKILE 875

Db  2802 AEKKHQALLLDTNKQYEVEIQT--YREKLTSKE-ECLSSQKLEIDLLKSSKEELNNSLK 2857

Qy   876 LQMQSIMEKKMKTK------RNIAKHYKNPSLITFLCKNCSVLACSGEDIHVIEK--- 925

Db  2858 ATTQ-ILEELKKTKMDNLKYVNQLKKENERAQGKMKLLIKSCKQLE---EEKEILQKELS 2913

Qy   926 -------MHHVNMTPBFKELYIVRENKA-----LQKKCADYQINGEIICK 963

Db  2914 QLQAAQEKQKQTGTVMDTKVDELITTEIKELKETLEEKTKEADEYLDKYCS-LLISHEKLEK 2972

Qy   964 CGQAWGTMVH 974

Db  2973 AKEMLETQVAH 2983
```

```
Query Match      3.5%;   Score 188.5;  DB 2;   Length 1388;
Best Local Similarity   18.2%;   Pred. No. 8.3e-07;
Matches  204;   Conservative 195;   Mismatches 399;   Indels 321;   Gaps  57;

Qy   71 WHLGWTREFVEALRRTGSPLAARYMNPEL-TDLPSPSFENAHDEYLQLLNLLQPTLVDKL 129
           ::|||   : |::  |  ||:|::      :|||  |:|:|  :
Db  359 WH--W----DNIRETAAPVV----PELSSDIDSSNFDDIEDD----KGDVETFPIPKA 402

Qy  130 LVRDVLD---KCMEELLITEDRNRIAAAENGNESGVRELLKRIVQKENWFSAFLNVL 185
            |:|::|     |::|   ::| :||  :     ||  |:|:||||:||| :
Db  403 FVGNQLPFIGFTYYRENLLLSDSP---SCRENDSIQSRKNEESQEIQKK------LYTL 452

Qy  186 RQTGNNEL-VQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMENNS 244
             |::|:|  |::|   :: |::   : :  :|      |::   :     |:||:  ||| :
Db  453 EEHLSNEMQAKEELEQKCKSVNTRLEKTAK---ELEEEI---TLRKSVESALRQLEREK 505

Qy  245 SESSFADSSVVSESDTSL----AEGSVSCLDESL------GHNSNMGSDS-GTMGSDSD 292
           ||  :  |:|::|  |                   :::         ||:: |   :       ::
```

; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 296
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 064987CD1
US-09-976-594-296

Query Match   3.5%;   Score 188.5;   DB 4;   Length 1388;
Best Local Similarity   18.2%;   Pred. No. 8.3e-07;
Matches   204; Conservative   195; Mismatches   399; Indels 321; Gaps   57;

Qy   71 WHLGWTREFVEALRRTGSPLAARYMNPEL-TDLPSPSFENAHDEYLQLLNLLQPTLVDKL 129
Db   359 WH--W---DNIRETAAPVV---PELSSDIDSSNFDDIEDD---KGDVETFPIPKA 402

Qy   130 LVRDVLD---KCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAFLNVL 185
Db   403 FVGNQLPFIGFTYYRENLLLSDSP--SCRENDSIQSRKNEESQEIQKK---LYTL 452

Qy   186 RQTGNNEL-VQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMENNS 244
Db   453 EEHLSNEMQAKEELEQKCKCKSVNTLEKTAK---ELEEEI---TLRKSVESALRQLEREK 505

Qy   245 SESSFADSSVVSESDTSL------AEGSVSCLDESL-------GHNSNMGSDS-GTMGSDSD 292
Db   506 ALLQHKNABYQRKADHEADKKRNLENDVNSLKDQLEDLKKGNQNSQISTEKVNQLQRQLD 565

Qy   293 EENVAARASPEPELQLRPYQMEVAQPA-------LEGKNIICLPTGSGKTRVAVYIA 343
Db   566 ETNALLRTESDTAARLRKTQAESSKQIQQLESNNRDLQDKNCL--LETAKLKLEKEFINL 623

Qy   344 KDHLDKKKASEPGKVIVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEV 403
Db   624 QSALESERRDRTHGSEII---------------NDLQG------RICGLEED 654

Qy   404 VKSCDIIISTAQI---LENSLLNLENGEDAGVQLS-DFSLIII------DECHH 447
Db   655 LKNGKILLAKVELEKRQLQERFTDLEK-EKSNMEIDMTYQLKVIQQSLEQEEAEHKATKA 713

Qy   448 --TNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQILGLTASPGVGGATKQAKAEEH 505
Db   714 RLADKNKIYESI-----BEAKSEAMKEMEKKLLE 753

Qy   506 IL---KLCANLDAFTIKTVKENLDQLKNQIQEPCKKPAIADATREDPFKEKLLEIMTRIQ 562
Db   754 LLEAEKRCSLLDC---DLKQSQQKINELLKQ--KDVLNED-VRNLTLKIEQETQ 801

Qy   563 TYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEAL---------QI 613
Db   802 KRC--LTQNDLKMQTQQVNTLKMSEKQLKQENNHLMEMKMNLEKQNAELRKERQDADGQM 859

Qy   614 NDTIRMIDAYTHLETFYN---------KFAVIEDDSD----------EEKDK 646
Db   860 KELQDQLEAEQYFSTLYKTQVRELKEECEEKTKLGKELQQKKQELQDERDSLAAQLEITL 919

Qy   647 ----------EGGDDEYCDGDED---EDDLKKPL---KLDETDRFLMTLFFENNKML 687
Db   920 TKADSEQLARSIAEEQYSDLEKEKIMKELEIKEMMARHKQELTEKDATIASL-EETNRTL 978

Qy   688 KRLAENPEYENEKLTKLRNTIMEQYTRTEE---SARGIIFTKTRQSAYALSQWITEN--- 741
Db   979 TSDVANLANEKEELNNKLKDVQEQLSRLKDEEISAAAI---------KAQFEKQLLTERTLK 1031

Qy   742 --EKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLL------- 786
Db   1032 TQAVNKLAEI-MNRKEPVKRGNDTDVR-RKEKENRKLHMELKSEREKLTQQMIKYQKELN 1089

RESULT 10
US-09-976-594-296
; Sequence 296, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594

```
Qy   327 ---ICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIVLVNKVLLVEQLFRKEFQPFLKK 383
                 :.: ||||                                  ||  :   :  :::
Db   810 ALPVCPITQATSCRAVTGLQQEH-----DKQKAEE-----LKQQVDELTAANRKAEQDMREL 861

Qy   384 WYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLENGEDAGVQLDFSLIID 443
                                                                      ::
Db   862 TYELNAL---QLEKTSSE-----EKARLLKDKL-------------------------D 887

Qy   444 ECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQILGLTASPGVGGATKQAKAE 503
                               ::    :: ||  | ||             ::  :  :: :  ::
Db   888 ET-------NNTLR----CLKLELERKDQAEKGYSQQLRELGRQLNQTTG-----KAE 929

Qy   504 EHILKLCANLDAFTIKTVK------ENLDQLKNQIOEPCKKFAIADATREDPFK---- 551
                                        |||:|.:| :              ::
Db   930 E------AMQEASDLKKIKRNYQLELESLNHEKGKLQREVDRITRAHAVAEKNIQHLNSQ 983

Qy   552 ------EKLLEIMTRIQTYCQWSPMSDFGTQPYEQWAIQMEK--KAAKKGN------- 594
                 :: :::                  .:  :  ::     :  :
Db   984 IHSFRDEKELE---RLQ-ICQRK--SDHLKEQPEKSHEQLLQNIKAEKENNDKIQRLNEE 1037

Qy   595 -RKERVCAEHL--------RKYN----------EALQI------ 613
                |:: :             ||:             ||:
Db  1038 LEKSNECAEMLKQKVEELTRQNNETKLMMQRIQAESENIVLEKQTIQQRCEALKIQADGF 1097

Qy   614 NDTIRMIDAYTHLETFYNEEKDKFAVIEDDSDEGGD-----DEYCDGDEDEDDLKKPLK 668
                                  :.  ||                                :  | ||:
Db  1098 KDQLRSTNEHLHKQTKTEQDFQRKIKCLEEDLAKSQNLVSEFKQKCD--------QQNII 1149

Qy   669 LDETDRFLMTLFPENN--KMLKRLAENP---------EYENEKLTKLRN---TIMEQYT 713
                 :                :  :  |:     : |||:                   .: : : ::
Db  1150 IQNTKKEVRNLNAELNASKEEKRRGEQKVQLQQAQVQELNNRLKKVQDELHLHLKTIEEQMT 1209

Qy   714 R-----TEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMT 767
                 |                |||
Db  1210 HRKMVLFQEESG-----KFKQSA------EEFRK---KMEKLM-----ESKVIT 1244

Qy   768 QNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADE 827
                                                                     |||
Db  1245 END----ISGIRLDFVSLQQENSRAQENAKLCETNI--------KELERQLQQYREQMQQ 1292

Qy   828 STYVLVAH-SGSGVIEHETVNDFREKMMYKAIHCVQNMKP-------EEYAHKILELQMQS 880
                  ::    |    :|| ||  |||
Db  1293 GQHMEANHYQKCQKLEDELIIAQKRE---------VENLKQKMDQQIKEHEHQLVLLQCEI 1343

Qy   881 IMEKKMKTKRNIAKGHYKNNPSLIITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYI 940
                                                                                 :.:  ||
Db  1344 -------QKKSTAKDCTFKPDF-----EMTVKECQHSGELSSRNTGHLHPTRSPLLRW 1390

Qy   941 VRENKALQKK 950
                 ::| : |||
Db  1391 TQEFQPLEEK 1400
```

RESULT 12
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SEQ ID NO 8
; SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 1786

```
Qy   787 -IATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARA-----DESTYVLVAHSGSGV 840
                 |||                                                  :::    ||
Db  1090 EMQAQAEE-----SQIRIELQMTLDSKDSDEQLRSQLQALHIGLDSSSI-----GSGP 1139

Qy   841 IEHETVNDFREKMM--YKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKN 898
                 :: : ::  :
Db  1140 GDAEADDGFPESRLEGWLSLPVRNNTKKFGWVKKYVIVSSKKILFYDSEQDKE-----QS 1194

Qy   899 NPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKEL---YIVRENKALQKKCADY 954
                 ||:  :                         |:   ::      :: :
Db  1195 NPYMVLDIDKLFHVRPVTQTDYY------RADAKEIPRIFQIIYANEGESKKEQEF 1244

Qy   955 QIN--GE---IICKCGQAWGTMVVH------------------------------ 974
                  :    ::
Db  1245 PVEPVGEKSNYICHKGHEFIPTLYHFPTNCEACMKPLWHMFKPPPALECRRCHIKCHKDH 1304

Qy   975 ------KGLDLPCLKIRNFVVVFKN-----NSTKKQYKKWV 1004
                   : :                          ||:: :  ||
Db  1305 MDKKEEIIAPC-KVYYDISTAKNLLLLANSTEEQ-QKWV 1341
```

RESULT 11
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1
US-09-919-172-29

Query Match   3.5%; Score 187.5; DB 4; Length 2125;
Best Local Similarity 19.6%; Pred. No. 2e-06;
Matches 214; Conservative 164; Mismatches 375; Indels 337; Gaps 51;

```
Qy    41 EVKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRT----------- 86
                 :::
Db   468 ENQPENSKTLATQLNQQ--KMLVSEIEMKQSKMDECQKYAEQYSATVKDYELQTMTYRAM 525

Qy    87 ----GSPLAARYMNPELTDLPSPSFENAHDEYIQLLNLLQPTIVDKLLVRDVLDKCMEE 141
                     :                           :: :  :       : :  :::
Db   526 VDSQQKSPVKRRRMQSS-ADLIIQEFMDLRTRYTALVTIMTQYI--KFAGDSLKRLEEE 581

Qy   142 ELLTIED---------RNRIAAAENN--GNESGVRELLKRIVQKENWFSAFLNVLR 186
                 :::  ::                                :::
Db   582 EIKRCKETSEHGAYSDLLQRQKATVLENSKLTGKISELERMVAELKKQKSRVEEELPKVR 641

Qy   187 QTGNEL-----VQELT------GSDCSESNAEIENL-SQVDGPQVE-EQLLSTTVQPN 232
                  : ::        :::                               :  :
Db   642 EAAENELRKQQRNVEDISLQKIRAESEAKQYRRELETIVREKEAAEERELERVRQLTIEAE 701

Qy   233 ----LEKEVWGMENNSSESSFADSSV---VSESDTSLAEGSVSCLDESLGHNSNMGSDS 284
                     ::                                 : :  :
Db   702 AKRAAVEENLLNFRNQLEENTFTRRTLEDHLKRKDLSL------NDLEQQKNKLMEE 752

Qy   285 GTMGSDSDEENVAARASPEPELQLRPYQMEVAQPALEGKNII---------------- 326
                                         :::|                   :::
Db   753 LRRKRDNEEELLKLIKQMEKDL---AFQKQVAEKQLKEKQKIELEARRKITEIQYTCREN 809
```

```
;  TYPE: PRT
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8

Query Match           3.5%;  Score 186.5;  DB 3;  Length 1786;
Best Local Similarity   19.6%;  Pred. No. 1.9e-06;
Matches  201;  Conservative 152;  Mismatches 373;  Indels 301;  Gaps 44;

Qy    42  VKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTD  101
              ::|: :|   :    |: :|
Db   499  VEESVAENVEESVAENVEEIVAPTVEEIV----------APTVEEIVAPSVVE  541

Qy   102  LPSPSPENAHDEYLQLLNLLQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGN-  160
              :|| |   :       ||
Db   542  SVAPSVEESVEEN----------VEESVAENVEESVAENVEESVAENVEEI  590

Qy   161  -----ESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAE-------I  209
                 ||
Db   591  VAPTVEEIVAPTVEEIVAPSVVESVAPSVEESVEEN--VEESVAENVEESVAENVEESVA  648

Qy   210  ENLSQVDGPQVEEQLLST-----------TVQPNLEKEVWGMENNSSESSFADSSVVSES  258
              :|:|:  :                    :|
Db   649  ENVEEIVAPTVEEIVAPTVEEIVAPSVVESVESV------SVAENV  699

Qy   259  DTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDERNVAARASPPPELQLRPYQMEVAQP  318
              :|  |        :|:
Db   700  EESVAENV----------VEESVAENV---EESVAENVEEIVAPTVEEIVAPTVEEIVAP  745

Qy   319  ALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNKVLLVEQLFRKEFQ  378
              :                   :| |:
Db   746  SV-----VESVAPSVEESVEEN-----------EESVAENVEESVAENVEESVAENVEESVA  792

Qy   379  PPLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLENGEDAGVQLSDFS  438
              :                            ::
Db   793  P---------------TVEEIVAPSVEE-----SVAPSVEESV------AENVATNLSDNL  827

Qy   439  LIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLK-----KENKPVIPLPQILGLTASPGV  493
                 :::|                            ||
Db   828  L-----------------SNLLGGIETEEIKDSILNEIEEVKENVVTTILENVEETTAE---  869

Qy   494  GGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKKPAIADATREDPFKEK  553
                 :|               :|:          ||
Db   870  -SVTTFSNILEIQE-----NTITNDTIEEKLEELHENVLS-----AALENTQSEEEKKE  918

Qy   554  LLEIM-------TRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEH  603
              ||::|         :|                      ||
Db   919  VIDVIEEVKEEVATTLIETVEQAEEKSANTITEIFE----NLEENAVESNEN------VAEN  970

Qy   604  LRKYNEFALQINDTIRMIDAYTHL-ETFYNEKDKKFAVIEDDSDEGGDDEYCDG---  656
              ||:                    :
Db   971  LEKLNETV-FNTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKGIQENLLTGMFRSI  1029

Qy   657  ------DEDEDDLKKPLKLDETDRFLMTLFENNKMLKRLAENPEYENE-KLTKLRNTI  708
              :::: ||
Db  1030  ETSIVIQSEEKVDL-----------NENVVSSILDNIENMKEGLLNKLENI-----  1069

Qy   709  MEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKA--HHLIG---AGHSSE  762
              :: :|:|                           |:    :::||  ||
Db  1070  ---------SSTEGVQETVTEH---------VEQNVYVDVDVPAMKDQFLGIINEAGGLKE  1112

Qy   763  F----------------KPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKE  800
              | :|                       :||:    |:
Db  1113  MFFNLEDVFKSESDVITVEEIKDEPVQKEVKETVSIIEEMEEN--IVDVLSEEKEDLTD  1170

Qy   801  CNIVIRYGLVITNEIAMVQARGRARADESTYVLVAHSGSGVIHEHETVNDFRRKKMYKAIHC  860
              :   | |               :|                    |
Db  1171  ---------KMIDA--VERSIEI-----SSDSKEETESIKD-KEKDVSLVVEE  1206

Qy   861  VQNMKPEEYAHKILELQMQSIMEKKMK-----------TKRNIAKHYKNN  899
              ||:|        ||:                    :
Db  1207  VQDNDMDESVEKVLEL--KNMEEELMKDAVEINDITSKLIEETQELNEVEADLIKDMEKL  1264
```

```
Qy   900  PSLITFLCKNCSVLACSGEDI--HVIEKMHVNMT-PEFKELYIVRENKALQKKCADYQI  956
              |: :            :|:     :|  |:
Db  1265  KELEKALSEDSKEIIDAKDDTLEKVIEEEHDITTLDEVELKDVEEDKI--EKVSDLKD  1322

Qy   957  NGEIICK  963
              |||:||:
Db  1323  LEEDILK  1329
```

RESULT 13
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match   3.4%; Score 182.5; DB 4; Length 2954;
Best Local Similarity   17.9%; Pred. No. 9.2e-06;
Matches 182; Conservative 187; Mismatches 366; Indels 283; Gaps 42;

```
Qy     6  STDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVELLLST  65
              |                :::  ::  :
Db   564  SKEEEKKELVQSFELKI-----------------AELEEQL------SVKAKNLEMVTNS  600

Qy    66  LEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFE----NAHDEYLQLLNL  120
              |                           ||:
Db   601  REHSINAEVQTDVEKEVVRKEMSVLGDSGYNASNSDLQDSSVDGKRLSSSHDECIEHRKM  660

Qy   121  LQPTLVD-------------------KLLVRDVLDKCMEEELLTIBDRN---RIAAAENN--  158
              ||:                     |:
Db   661  LEQKIVDLEEFIENLNKKSENDKQKSSEQDFMESIQLCEAIMAEKANALEEALMRDNFD  720

Qy   159  ---GNESGVREL--LKRIV--QKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEI  209
              :                   :||
Db   721  NIILENETLKREIADLERSLKENQETNEFEILKEKETQKEHEAQLIHEIGSLKKLVENAEM  780

Qy   210  ENLSQVDGPQVEEQLL-STTVQ-------PNLEKEV---------WGMENNSSESSFADS  252
              ||:| |            :|:  :                  |
Db   781  ... (continues)
```

Query Match   3.4%; Score 179; DB 4; Length 666;
Best Local Similarity   19.5%; Pred. No. 1.7e-06;
Matches 137; Conservative 114; Mismatches 230; Indels 220; Gaps 35;

```
Qy     253  SVVSEDTSLAEGSVSC------------LDESLGHNSNMGSDSGTMGS------DSDENV-    296
                 :::|                   ||:|  .: |.   :              ::::
Db     321  EGKNIICL-PTGSGKIRVAVYIAKDHLDKKKKASEPGK-VIVLVNKVLLVEQLFR--KE   376

Qy     297  AARASPEPELQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDK---KKKA    353
                  :  :                  || ||:||||||||||||||||||||
Db      37  EGKRHQTLLGATGTGKFTMSNVIK-------------EVGKPTLIIAHNKTLAGQLYSEFKE  86

Qy     354  SEPGKVI--------VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVK    405
                :: .|:|          :|||||||||||||||||||:|::
Db     377  FQPPLKKWYRVIGLSGDTQLKISFPEV-VKSCDIIISTAQILENSLLNLENGEDAGVQLS   435

Qy     406  SCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQK    465
                |||||||||||||||||||||||||||.   ||:||||||||||||:|||:||||||
Db      87  FFPENRVEFV-----SYYDYQPEAVVPSTDTFIEKDASINDEIDQLRHSATSSLFER     140

Qy     406  SCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQK    465
Db     436  DFSLII--IDECHHTNKEAVYNNIM----------RHYLMQKL------KNN--         469

Qy     466  LKNNRLKK--ENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKE    523
                | :|::|    ||::|:|||||||||||||||||||||||||||:|||||
Db     141  DDVIIIASVSCIYGLGNPEEYKNLVVSRVGMERSELLRKLVDVQYSRNDIDFQRGTF     200

Qy     524  NLDQLKNQIQEPCKKFAIADATREDPFKEKLLE------IMTRIQTYCQMSPMSDFGTQPYE   579
                           :::|                        ||||||||||||||||||
Db     470  RLKKENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAF--TIKTVKENLDQ   527

Qy     524  NLDQLKNQIQEPCKKFAIADATREDPFKEKLLE
Db     201  RVRGDVVEIFP-----------ASREE----MCIRVEFFGDEIDRIRE-VNY     236

Db     528  LKNQIQEPCKKFAIADA----TREDPFK-----------EKLLE                 556

Db     237  LTGEVIREREHFTIFPASHFVTREEKMKVAIERIEKELEERLKLRDENKLLEAQRLEQR   296

Qy     580  QWAIQMEKKAAKGNRKERVCAEHLRKYNEALQINDTIRMIDAYTH-------LETFYNE    632
Db     557  --------------IMTRIQTYCQMSPMSDFGTQPY-------------EQWAIQMEK-----KA  589

Qy     633  EKDK------------KFAVIEDDSDEGGDBEYCDGDEDEDLKKPLKLDETDRFLMTLFFENNKM   686
Db     297  TNYDLEMVMREMGFCSGIENYSVHLTLRPLGSTPYTLLDYFGDDWLVMIDESHVTLPQIRG   356

Qy     687  LKEQVIDLNTQLQSLQAQSIEKSDLQKPKQDLEEGEVKLLLEME----LLKGHLTDSQLS   744
Db     590  AKKGNR-KERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEKDKKFAVIEDDSDEG   648

Qy     745  ABVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFR------------------KEAGIKVNYL-    779
Db     357  MYNGDRARKQVLIDHGFRLPSALD-NRPLK----FEEF--EEKTKQLVYV----SATP   403

Db     649  GDDEYCDGDEDEDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTI   708

Db     404  G----------------PYELEHTDEMVEQIIRPTGLLDPKIDVRPT-ENQ------IDDLL   442

Qy     780  TGKINLLIATTVAEEGLDIK-ECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGS    838
Db     709  MEQYTRTBESARGIIFTKTRQSAYALSQWITENEKFABVGVKAHHLIGAGHSSEFKPMTQ   768

Qy     781  YNQNLEEDLETKLLKEQEIQLAELRKRADNLQKKVRNFDLSVSMGDSEKLCEEIFQLK    840
Db     769  NEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVI-------RYGLVTNEIAMVQARGR-   822

Qy     839  GVIEHETVNDFREKMMYKAIHCVQNMK----------PEEYAHKILELQMQSIMEKKMK    887
Db     443  SEIQDRVDKDERVLVTTLTKKMSEDLTTYM------KEAGIKVNYL-----HSEIKTL-   489

Qy     841  QSLSDAEAVTRDAQKECSFLRSENLELKEKMEDTSNWYNQKEKAASLFEKQLETEKSNYK   900
Db     490  -ERIEIIRDLRMGTYDAIVGINLLREGIDIPEVSLVVILDADKEGFLRSDRSLIQTIGRA   548

Qy     901  KMEADLQKELQSAFNEINYLNGLLAGK---------------VPRDLLSRVELEKKV    942
Db     823  ARADESTYVL-------------------------VAHSGSGVIEHETVNDFREKMMYKAIH   859

Qy     943  SEFSKQLEKALEEKNALENEVTCLSEY--KFLP-----------NEVECLKNQISK    985
Db     549  ARNDKGEVIMYADKITDSMQYAIDETQRRREIQIAHNKEHGITPKTIN--------KKIH   600

Qy    1045  MHDDLFEKYIRNK--------------SEAED-LLREMENLKG-TMESVEV    1079
Db     528  LKNQIQEPCKKFAIADA----TREDPFK

Qy    1080  KIADTKHELEE------TIRD------KEQLLHEKKYFFQAMQTIFPITPLSD---------   1120
Db     860  CVQNM---KPEEYAHKILELQMQSIMEKKMKTKRNIAKHYK  897

Qy    1121  ----SLPPSKLVEGNSQ-------------DPIEINDYHNLIALATERNNIMVCLETERNS   1164
Db     601  DVISATVESDETNQQQQTELPKKWTKKERQKTIENIEKEMK  641

Qy    1165  LKEQVIDLNTQLQSLQAQSIEKSDLQKPKQDLEEGEVKLLLEME----LLKGHLTDSQLS   1220

Qy    1221  IEKL-----QLENLEVTEKLQTLQEEMKNITIERNELQTNFEDLKAEHDSLKQDLSEN---   1273

Qy    1274  IEQSIETQDELRAAQEE-------LREQKLVDSFRQQLLDCSVGISSPNHDAVANQEKVS   1327

Qy    1328  LGEVNSLQSEMLRGERDELQTSCKALV-------SELELLRAHVKSVEGENLETTKKLNG-   1380

Qy    1381  --LEKEILGKSEESEVLKSM--LENLKEDNNKLKEQAEEYSSKENQFSLEEVFSGSGQK   1434
```

```
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/685,576
        FILING DATE: 24-JUL-1996
        CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: JP 7-325129
        FILING DATE: 20-NOV-1995
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: JP 8-17150
        FILING DATE: 05-JAN-1996
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: JP 8-131206
        FILING DATE: 26-APR-1996
    ATTORNEY/AGENT INFORMATION:
        NAME: Bent, Stephen A.
        REGISTRATION NUMBER: 29,768
        REFERENCE/DOCKET NUMBER: 16887/843
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (202)672-5300
        TELEFAX: (202)672-5399
        TELEX: 904116
US-08-685-576-1
INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1388 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
    MOLECULE TYPE: protein

Query Match       3.3%;   Score 174;  DB 2;  Length 1388;
Best Local Similarity  18.5%;  Pred. No. 1.5e-05;
Matches  207;  Conservative 186;  Mismatches 390;  Indels  334;  Gaps  58;

Qy    83 LRRTGSPLAARYMNPEL-TDLPSPSFENAHDEYLQLLNLLQPTLV------------D 127
            ||||| :|: ::|   ::|     ||: ::|::|:
Db   364 IRETAAPVV---PELSSDIDSSNFDDIEDDKGDVETFPIPKAFVGNQLPFIGFTYYRE 418

Qy   128 KLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAFLNVLRQ 187
            ||:||  :                ::|    :
Db   419 NLLLSD-SPSCKEND--SIQSRK-----------NEES-QEIQKKLYTLEEHLSTEIQ--- 461

Qy   188 TGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMENNSSES 247
            :||   :|:               ::|: |:  :|:|: |:
Db   462 -AKEELEQK-----CKSVNTRLEKVAK-----ELEEEI---TLRKNVESTLRQLEREKALL 508

Qy   248 SFADSSVVSESDTSL-----AEGSVSCLDESL------GHNSNMGSDS-GTMGSDSDEEN 295
            :: :  :: | : :     | |::  |: :        |||:  |:| |  :|:|:|:|
Db   509 QHKNAEBYQRKADHEADKKRNLENDVNSLKDQLEDLKRKRQNSQISTEKVNQLQRQLDETN 568

Qy   296 VAARASEPELQLRPYQMEVAQPA----------LEGKNIIICLPTGSGKTRVAVYIAKDH 346
            :|  |:||| :|     :||         ||:||     |: | ::
Db   569 ALLRTESDTAARLRKTQAESSKQIQQLESNNRDLQDKN--CI------------LETAKLK 615

Qy   347 LDKKKKASEPGKVIVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKIS-FPEVVK 405
            |:|     :   :|:|: |||  : : |  |:|:   :   |:|  :|  |  ::|:|:|
Db   616 LEKE-------FINLQSVLESERRDRTHGSEII-----------NDLQGRISGLEBDVK 656

Qy   406 SCDIIISTAQI---LENSLLNLENGEDAGVQLS-DFSLIII------------DECHI----- 447
            :|:||:|:::|    |::|:|: || ||:  |  |:::::            |||:
Db   657 NGKILLAKVELEKRQLQERFTDLEK-EKNNMEIDMTYQLKVIQQSLEQEETEHKATKARL 715

Qy   448 TNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIIGLTASPGVGGATKQAKAEEHIL 507
            :|:|  :  :    ::|:|        :||::       ::|: ::||| :|||::||:|
Db   716 ADKNKIYESI--------EEAKSEAMKEMEKKLSE----------ERTLKQKVENLLL 755

Qy   508 --KLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFFKEKLLEIMTRIQTY 564
            ||:  |  :||::|||   :||:|:|| :  | ::||::|||||||:|:||||::||||
Db   756 EABKRCSIIDC----------DLKQSQQKINELLKQ---KDVLNED-VRNLTLKIEQETQKR 803
```

```
Qy   565 CQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEAL-----------QIND 615
            ::  |   ::|: :: :::|   |:|::|:||::  ::||:|:||::           |:|:
Db   804 C--LTQNDLRMQTQQVNTLKMSEKQLKQENNHLLEMKQMSLEKQNAELRKERQDADGQMKE 861

Qy   616 TIRMIDAYTHLETFYN-------------EKKDK-------KPAVIEDDSD---------- 646
            |:|::  :| ||::::             ||:|:       ||::|:|:||
Db   862 LQDQLEAEQYFSTLYKTQVRELKEECEEKTLCKELQQKKQELQDERDSLAAQLEITLTK 921

Qy   647 -------EGGDDEYCDGDED---EDDLKKPL---KLDETDRPLMTLFFENNKMLKR 689
                   |  ::|: |::|:   ||:|:|:|   |::|: |: |: ::||::|||:
Db   922 ADSEQLARSIAEEQYSDLEKEKIMKELEIKEMMARHKQELTEKDATIASL-EETNRTLTS 980

Qy   690 LAENPEYENEKLTKLRNTIMEQYTRTEE---SARGIIFTKTRQSAYALSQWITEN------ 741
            :|:|   ::||||:|:|   |:|:|||:    |:|:|||||:|: :|:|||:|:|
Db   981 DVANLANEKEELNNKLKEAQEQLSRLKDEISAAAI------KAQFEKQLLFERTLKTQ 1033

Qy   742 --EKFAEVGVKAHHLLGAGHSSEFKPMTQNEQKEVLSKFRTGKINLL----------I 787
            | |:||: | : :|:   |:: | |::  |::|:|:|:::|:|:|::           |
Db  1034 AVNKLAEI-MNRKEPVKRGNDTDVR-RKEKENRKLHMELKSEREKLTQQMIKYQKELNEM 1091

Qy   788 ATTVABEGLDIKECNIVIRYGLVTNEIAMVQARGRARA-----DESTYLVAHSGSGVIE 842
            |:::|:|  :|:|  :  :::    ::|::|:|  :|:      :: ::|:  ||:|  |
Db  1092 QAQIAEE----SQIRIELQMTLDSKDSDIEQLRSQLQALHIGLDSSSI-----GSGPGD 1141

Qy   843 HETVNDFREKMM--YKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
            :|:: |:  :::  ||::   ::  ::| ::|::||:::::||:|:::|:| ::||:||:|
Db  1142 TEADDGFPESRLEGWLSLPVRNNTKKFGWVKKYVIVSSKKILFYDSEQDKE------QSNP 1196

Qy   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKEL------YIVRENKALQKKCADYQI 956
            ::::|:|| |:||:| |:|||:|:| :||::|||||::|      |::||:|:|:|:|||:|:
Db  1197 YMVLDIDKLFHVRPVTQTDVY-------------RADAKEIPRIFQILYANEGESKKQEFPV 1246

Qy   957 N--GB--IICKCGQAWGTMMVH---------------------------------------- 974
            :  |:  |:|:|| ::|:::::
Db  1247 EPVGEKSNYICHKGHEFIPTLYHFPTNCEACMKPLWHMFKPPPALECRRCHIKCHKDHMD 1306

Qy   975 --KGLDLPCLKIRNFVVVFKN-----NSTKKQYKKWV 1004
            | |::::||:|:|::|:|::|     |||::|||||
Db  1307 KKEEIIAPC-KVYYDISSAKNLLLLANSTEEQ-QKWV 1341
```

Search completed: June 2, 2004, 19:13:05
Job time : 29 secs

```
OM protein - protein search, using sw model

                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:       June  2, 2004, 19:11:53 ; Search time 54 Seconds
                                         (without alignments)
                                         5340.225 Million cell updates/sec

Title:        US-09-515-363C-2
Perfect score: 5311
Sequence:     1 MSNGYSTDENFRYLISCFRA........LPTIFPNLDYSECCLFSDED 1025

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:     1155919 seqs, 281138677 residues

Total number of hits satisfying chosen parameters:    1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Published_Applications_AA:*
    1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
    2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
    3: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
    4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
    5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
    6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
    7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
    8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
    9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
   10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
   11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
   12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
   13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
   14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
   15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
   16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
   17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
   18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	% Match	Query Match	Length	DB	ID	Description
1	5311	100.0	1025	14	US-10-055-475-2	Sequence 2, Appli
2	5311	100.0	1025	14	US-10-228-897-2	Sequence 2, Appli
3	5307	99.9	1025	14	US-10-055-475-9	Sequence 9, Appli
4	5299	99.8	1025	14	US-10-055-475-7	Sequence 7, Appli
5	5285	99.5	1025	16	US-10-275-822A-2	Sequence 2, Appli
6	2456	46.2	468	15	US-10-094-749-2467	Sequence 2467, Ap
7	1087.5	20.5	925	16	US-10-408-765A-2031	Sequence 2031, Ap
8	1076	20.3	486	14	US-10-055-475-12	Sequence 12, Appl
9	1040.5	19.6	487	15	US-10-108-260A-4515	Sequence 4515, Ap
10	788.5	14.8	514	14	US-10-055-475-13	Sequence 13, Appl
11	783	14.7	166	9	US-09-864-761-45372	Sequence 45372, A
12	782	14.7	154	14	US-10-106-698-5088	Sequence 5088, Ap
13	782	14.7	154	15	US-10-264-049-2244	Sequence 2244, Ap
14	617	11.6	416	14	US-10-055-475-14	Sequence 14, Appl
15	563.5	10.6	447	15	US-10-264-049-2303	Sequence 2303, Ap
16	488	9.2	92	9	US-09-864-761-33849	Sequence 33849, A
17	453	8.5	752	15	US-10-369-493-21554	Sequence 21554, A
18	441.5	8.3	650	15	US-10-369-493-1338	Sequence 1338, Ap
19	441.5	8.3	650	15	US-10-369-493-20401	Sequence 20401, A
20	417	7.9	741	15	US-10-369-493-21360	Sequence 21360, A
21	396.5	7.5	485	14	US-10-055-475-17	Sequence 17, Appl
22	384	7.2	549	14	US-10-055-475-16	Sequence 16, Appl
23	357	6.7	71	9	US-09-864-761-47821	Sequence 47821, A
24	342	6.4	738	15	US-10-369-493-1227	Sequence 1227, Ap
25	340	6.4	503	14	US-10-055-475-15	Sequence 15, Appl
26	332	6.3	1924	9	US-09-866-557A-2	Sequence 2, Appli
27	332	6.3	1924	11	US-09-858-862-2	Sequence 2, Appli
28	332	6.3	1924	14	US-10-055-797-2	Sequence 2, Appli
29	332	6.3	1924	16	US-10-350-798-2	Sequence 2, Appli
30	298	5.6	55	13	US-09-864-761-48023	Sequence 48023, A
31	286	5.4	503	13	US-10-027-806-66	Sequence 66, Appl
32	286	5.4	503	13	US-10-034-623-66	Sequence 66, Appl
33	286	5.4	503	14	US-10-027-801-66	Sequence 66, Appl
34	286	5.4	503	14	US-10-029-120-66	Sequence 66, Appl
35	284	5.3	502	13	US-10-027-806-34	Sequence 34, Appl
36	284	5.3	502	14	US-10-034-623-34	Sequence 34, Appl
37	284	5.3	502	14	US-10-027-801-34	Sequence 34, Appl
38	284	5.3	502	14	US-10-029-120-34	Sequence 34, Appl
39	283	5.3	56	15	US-09-864-761-46695	Sequence 46695, A
40	277	5.2	669	15	US-10-104-047-3611	Sequence 3611, Ap
41	244	4.6	50	14	US-10-228-897-10	Sequence 10, Appl
42	214.5	4.0	1790	16	US-10-369-493-1586	Sequence 1586, Ap
43	205.5	3.9	1583	16	US-10-408-765A-1635	Sequence 1635, Ap
44	203	3.8	1905	15	US-10-259-194A-86	Sequence 86, Appl
45	202	3.8	3878	14	US-10-080-608A-11	Sequence 11, Appl

```
                            ALIGNMENTS

RESULT 1
US-10-055-475-2
; Sequence 2, Application US/10055475
; Publication No.US20030022855A1
; GENERAL INFORMATION:
;  APPLICANT: Fisher, Paul B.
;  APPLICANT: Kang, Dong-Chul
;  APPLICANT: Gopalkrishnan, Rahul V.
;  TITLE OF INVENTION: USE OF MDA-5 AS AN ANTIVIRAL AND
;  TITLE OF INVENTION: ANTIPROLIFERATIVE AGENT
;  FILE REFERENCE: A34614-A-PCT-USA-A (070050.1921)
;  CURRENT APPLICATION NUMBER: US/10/055,475
;  CURRENT FILING DATE: 2002-01-22
;  PRIOR APPLICATION NUMBER: PCT/US01/06960
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: 09/515,363
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;  LENGTH: 1025
;  TYPE: PRT
;  ORGANISM: homo sapiens
US-10-055-475-2

  Query Match             100.0%;  Score 5311;  DB 14;  Length 1025;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1025;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy          1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy         61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy        121 LQPTLVDKLLVRDVLDKCMEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
```

```
;   TITLE OF INVENTION:  GENE-5 AND PROMOTER AND USES THEREOF
;   FILE REFERENCE: A34614-A-PCT-USA (070050.2121)
;   CURRENT APPLICATION NUMBER: US/10/228,897
;   CURRENT FILING DATE:  2002-08-26
;   PRIOR APPLICATION NUMBER: PCT/US01/06960
;   PRIOR FILING DATE:  2001-02-28
;   PRIOR APPLICATION NUMBER: 09/515,363
;   PRIOR FILING DATE:  2000-02-29
;   NUMBER OF SEQ ID NOS: 25
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1025
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-228-897-2

Query Match         100.0%;  Score 5311;  DB 14;  Length 1025;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1025;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
Db    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
Db   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
Db  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
Db  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDSDENVAARA 300
Db  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDSDENVAARA 300

Qy  301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360
Db  301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360

Qy  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
Db  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy  421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480
Db  421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480

Qy  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540
Db  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540

Qy  541 IADATREDPFFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
Db  541 IADATREDPFFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600

Qy  601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEBKDKKFAVIEDDSDEGGDEYCDGDEDE 660
Db  601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEBKDKKFAVIEDDSDEGGDEYCDGDEDE 660

Qy  661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720
Db  661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720

Qy  721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780
Db  721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780

Qy  781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
```

Left-column alignment continuation:

```
Db  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
Db  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDSDENVAARA 300
Db  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDSDENVAARA 300

Qy  301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360
Db  301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360

Qy  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
Db  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy  421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480
Db  421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480

Qy  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540
Db  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540

Qy  541 IADATREDPFFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
Db  541 IADATREDPFFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600

Qy  601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEBKDKKFAVIEDDSDEGGDEYCDGDEDE 660
Db  601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEBKDKKFAVIEDDSDEGGDEYCDGDEDE 660

Qy  661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720
Db  661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720

Qy  721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780
Db  721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780

Qy  781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
Db  781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840

Qy  841 IEHETVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
Db  841 IEHETVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900

Qy  901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960
Db  901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960

Qy  961 ICKCGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKMVELPITFPNLDYSECCL 1020
Db  961 ICKCGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKMVELPITFPNLDYSECCL 1020

Qy 1021 FSDED 1025
Db 1021 FSDED 1025
```

```
RESULT 2
US-10-228-897-2
; Sequence 2, Application US/10228897
; Publication No. US20030092043A1
; GENERAL INFORMATION:
;   APPLICANT: Fisher, Paul B.
;   APPLICANT: Kang, Dong-Chul
;   APPLICANT: Gopalkrishnan, Rahul V.
;   TITLE OF INVENTION: MELANOMA DIFFERENTIATION ASSOCIATED
```

```
Db   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900

Qy   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960

Qy   961 ICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCL 1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCL 1020

Qy  1021 FSDED 1025
             |||||
Db  1021 FSDED 1025
```

RESULT 3
US-10-055-475-9

```
Query Match             99.9%;  Score 5307;  DB 14;  Length 1025;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1024;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

Qy     1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy    61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy   121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy   181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy   241 ENNSBSSSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ENNSBSSSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300

Qy   301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI 360

Qy   361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy   421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKQNNRLKKENKPVIP 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKQNNRLKKENKPVIP 480

Qy   481 LPQILGLTASPGVGGATKQAXAEEHILKLCANLDAFFIKTVKENLDQLKNQIQEPCKKFA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 LPQILGLTASPGVGGATKQAXAEEHILKLCANLDAFFIKTVKENLDQLKNQIQEPCKKFA 540

Qy   541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKERVC 600

Qy   601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660

Qy   661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720

Qy   721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780

Qy   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV 840

Qy   841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900

Qy   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960

Qy   961 ICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCL 1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCL 1020

Qy  1021 FSDED 1025
             |||||
Db  1021 FSDED 1025
```

RESULT 4
US-10-055-475-7

```
Query Match          99.8%;  Score 5299;  DB 14;  Length 1025;
Best Local Similarity 99.8%;  Pred. No. 0;
Matches 1023;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVVWGM 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300

Qy  301 SPEPELQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVVIAKDHLDKKKKASEPGKVI 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 SPEPELQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVVIAKDHLDKKKKASEPGKVI 360

Qy  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy  421 LLNLENGEDAGVQLSDFSLIIIDECHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 LLNLENGEDAGVQLSDFSLIIIDECHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480

Qy  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540

Qy  541 IADATREDPFFKEKLLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ::|||||
Db  541 IADATREDPFFKEKLLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKERVC 600

Qy  601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE 660

Qy  661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR 720

Qy  721 GIIFTKTRQSAYALSQWITENEKFABVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GIIFTKTRQSAYALSQWITENEKFABVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT 780

Qy  781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYYLVAHSGSGV 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYYLVAHSGSGV 840

Qy  841 IEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
        ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 IERETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900

Qy  901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI 960

Qy  961 ICKCCGQAWGTMVHKGLDLPCLKIRNFVVVFKNSTKKQYKKWVELPITFNLDYSECCLL 1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 ICKCCGQAWGTMVHKGLDLPCLKIRNFVVVFKNSTKKQYKKWVELPITFNLDYSECCLL 1020
```

```
Qy 1021 FSDED 1025
        |||||
Db 1021 FSDED 1025
```

RESULT 5
US-10-275-822A-2
; Sequence 2, Application US/10275822A
; Publication No. US2004008650A1
; GENERAL INFORMATION:
;   APPLICANT: Bahr, Georges
;   APPLICANT: Cocude, Cecile
;   APPLICANT: Capron, Andre
;   TITLE OF INVENTION: RH16 Polypeptide and its Fragments and Polynucleotides Encoding
;   TITLE OF INVENTION: said Polypeptides and Therapeutic Uses
;   FILE REFERENCE: 017753-170
;   CURRENT APPLICATION NUMBER: US/10/275,822A
;   CURRENT FILING DATE: 2002-11-12
;   PRIOR APPLICATION NUMBER: PCT/FR01/01441
;   PRIOR FILING DATE: 2001-05-11
;   PRIOR APPLICATION NUMBER: FR 00/06,030
;   PRIOR FILING DATE: 2000-05-11
;   NUMBER OF SEQ ID NOS: 56
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1025
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-275-822A-2

```
Query Match           99.5%;  Score 5285;  DB 16;  Length 1025;
Best Local Similarity 99.5%;  Pred. No. 0;
Matches 1020;  Conservative 2;  Mismatches 3;  Indels 0;  Gaps 0;

Qy    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180

Qy  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240

Qy  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA 300

Qy  301 SPEPELQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVVIAKDHLDKKKKASEPGKVI 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 SPEPELQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVVIAKDHLDKKKKASEPGKVI 360

Qy  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENS 420

Qy  421 LLNLENGEDAGVQLSDFSLIIIDECHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 LLNLENGEDAGVQLSDFSLIIIDECHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480

Qy  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA 540

Qy  541 IADATREDPFFKEKLLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  541 IADATREDPFFKEKLLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKESVC 600
```

```
Query Match             46.2%;  Score 2456;  DB 15;  Length 468;
Best Local Similarity   99.1%;  Pred. No. 9e-171;
Matches 464;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy   558  MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTI  617
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     1  MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEGNRKERVCAEHLRKYNEALQINDTI   60

Qy   618  RMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM  677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  RMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM  120

Qy   678  TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW  737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW  180

Qy   738  ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRTGKINLLIATTVAEEGLD  797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGRINLLIATTVAEEGLD  240

Qy   798  IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKA  857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIERETVNDFREKMMYKA  300

Qy   858  IHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG   917
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  IHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG   360

Qy   918  EDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMMVHKGL  977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  EDIHVIEKMHHVNMTPEFKELYIVRENKTLQKKCADYQINGEIICKCGQAWGTMMVHKGL  420

Qy   978  DLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED  1025
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   421  DLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED   468
```

RESULT 7
US-10-408-765A-2031
; Sequence 2031, Application US/10408765A
; Publication No. US2004010187A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2031
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2031

```
Query Match             20.5%;  Score 1087.5;  DB 16;  Length 925;
Best Local Similarity   31.5%;  Pred. No. 2.5e-70;
Matches 328;  Conservative 166;  Mismatches 361;  Indels 185;  Gaps  37;

Qy    12  RYLISCFRARVKMYIQVEPVLDYLT--FLPAEVKEQIQRTVATSGNMQAVELLLSTLEKG   69
          :  ::  :. : .  :  : ::  . .   ::: : :: ::: :    :..:: .:  :
Db     6  RRSLQAFQDYIRKTLDPTYILSYMAPWFREEV-QYIQAEKNNKGPMEARTLFLKFLLE-   63

Qy    70  VWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAH--DEYLQLLNLLQPTLVD  127
          :  :. ::.: .  : . .   : :.:. . :  ::..    : : ::. : :: . :
Db    64  LQEEGWFRGFLDALDHAG----YSGLYEAISWD-------FKKIEKLEEYRLLLKRLQPEFKT  116
```

RESULT 6
US-10-094-749-2467
; Sequence 2467, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2467
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2467

```
; APPLICANT: Gopalkrishnan, Rahul V.
; TITLE OF INVENTION: USE OF MDA-5 AS AN ANTIVIRAL AND
; TITLE OF INVENTION: ANTIPROLIFERATIVE AGENT
; FILE REFERENCE: A34614-A-PCT-USA-A (070050.1921)
; CURRENT APPLICATION NUMBER: US/10/055,475
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US01/06960
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/515,363
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 486
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-055-475-12

Query Match             20.3%;  Score 1076;  DB 14;  Length 486;
Best Local Similarity   44.3%;  Pred. No. 6.6e-70;
Matches  236;  Conservative   95;  Mismatches  154;  Indels   48;  Gaps   9;
```

```
Qy  306 LQLRPYQMEVAQPALEGKNIIICLPTSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK 365
            :::|| |   ||||||||| ||||| ||||||  ||| ||    : |   ||:|:|:
Db    1 MELRSYQWEVIMPALEGKNIIIWLPTGAGKTRAAAYVAKRHLE---TVDGAKVVVLVNR 56

Qy  366 VLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
        | |||| ||:||   |     |  | || | |: |      ||:::|:|  ||: : ::|
Db   57 VHLVTQ-HGEEFRRMLDGRWTVTTLSGDMGPRAGFGHLARCHDLLICTAELLQMALTSPE 115

Qy  426 NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIL 485
             |   : |||:||:|||||||| |: |||:||| || |||    |:     PLPQVL
Db  116 --EEEHVELTVFSLIVVDECHHTHKDIVYNVIMSQYLELKLQRAQ--------PLPQVL 164

Qy  486 GLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT 545
        |||||||:||| | :  |  :|:|||||||:  |   ::: |  :  |  ||::| :
Db  165 GLTASPGTGGASKLDGAINHVLQLCANLDTWCIMSPQNCCPQLQEHSQQPCKQYNLCHRR 224

Qy  546 REDPFKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL 604
         :||| : ||::| :|  :: :|| | ||||: |||  ::| :     ||::: |  :||
Db  225 SQDPFGDLLKKLMDQIHDHLEMPELSRKFGTQMYEQQQVVKLSEAAALAGLQEQRVYALHL 284

Qy  605 RKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDBGGDDEYCDGDEDEDDLK 664
        ::||:||  :||:| |||  : | :|:     :| ::  :               C
Db  285 RRYNDALLIHDTVRAVDALAALAQDFYHREHVTKTQIL---------------C------ 322

Qy  665 KPLKLDETDRFLMTLFPENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIF 724
              ||| | : | ||  :|: :|:|| | :|| ::|:|:|: :|||| |||||:||||||
Db  323 ------AERRLLALFDDRKNELAHLLATHGP-ENPKLEMLEKILQRQFS-SSNSPRGIIF 373

Qy  725 TKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKIN 784
        |:||||| :|  |  : :   :: | | |||||| ||:   |||::::||| || :| :|
Db  374 TRTRQSAHSLLLWLQQQGLQTVDIRAQLLIGAGNSSQSTHMTQRDQQEVIQKFQDGTLN 433

Qy  785 LLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSG 837
        || |: ||||||||  |:| |||| ||: :| ||||| |:|  : :|: ||
Db  434 LLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRAWADQSVVAFVATEG 486
```

```
Qy  128 KLIVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQ--KENWFSAFLNVL 185
        ::|    : ::: ::||:|::|  |  :   :  : :  ::     |    :|| : |
Db  117 RIIPTDIISD-LSECLINQECEILQICSTKGMAGAEKLVECLLRSDKENWPKTL---- 171

Qy  186 RQTGNNELVQELTGSDCGESNABIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMEN-- 242
          :: :|:::| :|  ||  : : : ::   ::     :    :  :  :   :
Db  172 ------KLALEKERNKFSELWIVEKGIKDVETEDLEDRKMETSDIQ------IFYQEDPEC 219

Qy  243 -NSSESSFADSSVVSESDTSLABGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARAS 301
          |::|:|  :|          ::                                  YS
Db  220 QNLSENSCPPSEV--SDTNL---------------------------------YS 239

Qy  302 PEPELQLRPYQMEVAQPALEGKNIIICLPTSGKTRVAVYIAKDHLDKKKKASEPGKVIV 361
             |     ||: |  |  ||||| ||| |:||| |  ||   ||    : |   :||:|
Db  240 P---FKPRNYQLELALPAMKGKNTIICAPTGCGKTFVSLLICEHHL-KKFPQGQKGKVVF 295

Qy  362 LVNKVLLVEQ--LFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILE 418
         :|: |::::  :| | |    :  ::|  :|| ::: |  :   :  :|||: |::||
Db  296 FANQIPVEQQKSVFSKYFE--RHGYRVTGISGATAENVPVEQIVENNDIIILLTPQILV 352

Qy  419 NSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPV 478
        | : :   :   : ::: ::|:  |||| |:: : |:|||: ||  || :
Db  353 N---NLKKGTIP--SLSIFTLMFDECHNTSKQHPYNMIMFNYLDQKLGGSS--------- 399

Qy  479 IPLPQIIGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKK 538
        :|||:||||||: ||   :|: :  :  : ||| |||: : |||:|  :|:: :  :|:|
Db  400 GPLPQVIGLTASVGVGDAKNTDEALDYICKLCASLDASVIATVKHNLEELEQVVYKPQKF 459

Qy  539 FAIADATREDPFKE-----------KLLEIMTRIQTYCQMSPMSDFGTQPYEQW 581
        | :  :    ||::           |: : | || :: :| :   ||||: |||
Db  460 FRKVESRISDKFKYIIAQLMRDTESLAKRICKDLENLSQION------REFGTQKYEQW 512

Qy  582 AIQMEKKA-----AKKGNRKERVC------AEHLRKYNEALQINDTIRMIDAYTHLETFYN 631
        : : :|:      ::|  :: : :      : :|:||||:|:  :| | | |  : |::
Db  513 IVTVQKACMVFQMPDKDEESRICKALFLYTSHLRKYNDALIISEHARMKDALDYLKDFFS 572

Qy  632 EEKDKKFPAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLA 691
        :    |:  ::                    :  :  :| ||:     ||:::::: :: :
Db  573 NVRAAGF---------DEIEQDL------TQR-----FEEKLQELESVS 601

Qy  692 ENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEK--FAEVGV 749
         :|  :|   |::| :   : : ::  : :  : :||:: :: ::|  :| :  | :||
Db  602 RDPSNENPKLEDLCFLIQEEYHLNPETIT-ILFVKTRALVDALKNWIEGNPKLSFLKPGI 660

Qy  750 KAHHLIGAGHSSEFKPMTQNEQKEVISKFR-TGKINLLIATTVAEEGLDIKECNIVIRYG 808
           |:|:| |:: :    |    :| :| :  :| ||:||||:|:|:|||:: |:|:: |:
Db  661 ---LITGRGKTNQNTGMTLPAQKCILDAFKASGDHNLIIATSVADEGIDIAQCNLVLLYE 716

Qy  809 LVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPEE 868
         | |:: ::|  |||||     | ::    :||:   :|:  ::|| ||: : : : |||
Db  717 YVGNVLKMIQTPRGRGRARGSKCFLLT-SNAGVIEKEQINMYKEKMMNDSILRLQTWDEAV 775

Qy  869 YAHKILELQMQSIMEKKMKTKRNIAKHY--KNNPSLITFLCKNCSVLACSGEDIHVIEKM 926
          :|| : : : :    :: :: ::: :     |:   |||: |:  ||  ::::|:|
Db  776 FREKILHIQTH---EKFIRDSQEKPKPVPDKENKKL---LCRKCKALACYTADVRVIEEC 829

Qy  927 HHVNMTPEFKELYIVRENKALQKKCADYQINGEIIC---KCGQAWGTMMVHKGLDLPCLK 983
         : : :  |:| ::   :  :    |   ::||::    |||:| | ||: || ::| :|
Db  830 HYTVLGDAFKECFVSRPHPK-PKQFSSFEKRAKIFCARQNCSHDWGIHVKYKTFEIPVIK 888

Qy  984 IRNFVVFKNNSTKKQYKKW 1003
        | ||||  ::: : |  |
Db  889 IESFVVEDIATGVQTLYSKW 908
```

```
RESULT 8
US-10-055-475-12
; Sequence 12, Application US/10055475
; Publication No. US20030022855A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Kang, Dong-Chul
```

```
RESULT 9
US-10-108-260A-4515
; Sequence 4515, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US2004005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4515
; LENGTH: 487
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4515

Query Match      19.6%;   Score 1040.5;  DB 15;  Length 487;
Best Local Similarity   41.1%;   Pred. No. 2.6e-67;
Matches  234;  Conservative  90;  Mismatches 161;  Indels  85;  Gaps  9;

Qy  306  LQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK  365
              ::|:|:||||  :| ::|:|||||||||||||||  ||||:||:||
Db    1  MELRSYQWEVIMPALEGKNIIIWLPTGAGKTRAAAYVAKRHLE---TVDGAKVVVLVNR   56

Qy  366  VLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE  425
              ::|:||  :||: ||||||||||||||||||||| ::
Db   57  VHLVTQ-HGEEFRRMLDGRWTVTTLSGDMGPRAGFGHLAR--------------------   95

Qy  426  NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIL  485
                                                            -------------
Db   96  --------------------CHHTHKDTVYNVIMSQYLELKLQRAQ----------PLPQVL  127

Qy  486  GLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT  545
              ||||||||||  :|:||::||:|||   :||   ||| ||| :  :
Db  128  GLTASPGTGGASKLDGAINHVLQLCANLDTWCIMSPQNCCPQLQEHSQQPCKQYNLCHRR  187

Qy  546  REDPPKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL  604
              ::|| |:| | :|| ||| || : ::  ||: : |::|  | ::|
Db  188  SQDPFGDLLKKLMDQIHDHLEMPELSRKFGTQMYEQQVVKLSEAAALAGLQEQRVYALHL  247

Qy  605  RKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLK  664
              |||: :  :||::||:||| ::|
Db  248  RRYNDALLIHDTVRAVDALAALQDFYHREHVTKTQIL----------------C-----  285

Qy  665  KPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIF  724
              :|  :|:|||: | | ||| ::: |||||||||||||
Db  286  -----AERRLLALFDDRKNELAHLATHGP-ENPKLEMLEKILQRQFS-SSNSPRGIIF   336

Qy  725  TKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKIN  784
              ||||||||:||||||||||||||:|:  | :  : ||| :||::|||| |||:||:|:|
Db  337  TRTRQSAHSLLLWLQQQGLQTVDIRAQLLIGAGNSSQSTHMTQRDQQEVIQKFQDGTLN  396

Qy  785  LLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHE  844
              ||:|||||||||| ||| |:|||||| ||||:|||||||||:| |  |   :
Db  397  LLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRARADQSVYAFVATEGSRELKRE  456

Qy  845  TVNDFREKMYKAIHCVQNMKPEEYAHKIL  874
              :| | ::| |:  :  |::| ::|:||:|
Db  457  LINEALETLMEQAVAAVQKMDQAEYQAKEL  486
```

```
RESULT 10
US-10-055-475-13
; Sequence 13, Application US/10055475
; Publication No. US20030022855A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Kang, Dong-Chul
; APPLICANT: Gopalkrishnan, Rahul V.
; TITLE OF INVENTION: USE OF MDA-5 AS AN ANTIVIRAL AND
; TITLE OF INVENTION: ANTIPROLIFERATIVE AGENT
; FILE REFERENCE: A34614-A-PCT-USA-A (070050.1921)
; CURRENT APPLICATION NUMBER: US/10/055,475
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US01/06960
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/515,363
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 514
; TYPE: PRT
; ORGANISM: sus scrofa
US-10-055-475-13
```

```
Query Match     14.8%;   Score 788.5;  DB 14;  Length 514;
Best Local Similarity   37.4%;   Pred. No. 7.3e-49;
Matches  210;  Conservative  89;  Mismatches 178;  Indels  85;  Gaps  16;

Qy  306  LQLRPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK  365
              ||:|:||  | ||: |||:|||  |  ||||:|:|| ||  ||| ||:||| ||||
Db    6  LKPRKYQLELALPAQNGKNTIICAPTGCGKTFVSLLICEHHL-KKFPRGRKGKVVFFAIQ   64

Qy  366  VLLVEQ--LFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLL  422
              : :||    |  : ::| ||||:|:|| | ::| ::  |:|:   :|||||||:||:||
Db   65  LPVYEQQKSVFSKHFE--RLGYKVAGISGATSDTVCVEQIVENSDIIILFPQILVNCLT  121

Qy  423  NLENGBDAGV--QLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP  480
                                                                     ----DS
Db  122  N-----GTIPSLSVFTLMIFDECHNTSKQHPYNVIMFSYLDRKLGGSS------------DS  166

Qy  481  LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA  540
              ||| ||||||||:|||:| | ::||||||||| :||:|| |:||| |  :
Db  167  LPQVIGLTASPGVGVGDAKNKAEATEYICKLCASLDTSVIATVRDNLELEEVVYKPQKFFR  226

Qy  541  IADATREDPPKEKLLEIMTRIQTYCQ------------MSPMSDFGTQPYEQWAIQM  585
              :   ||:| |:|| :::|  |||||              :  ||| |:|:||  ||
Db  227  KVELRTTDRFKCIISQLMHIESLAKSIFEELGTTILGGLFQIQNSNFGTQKYEQWITKV  286

Qy  586  EKKAA---KKGNRKERVC------AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKD  635
              :|:       |:|  :| C        || ||  | :||:||:||| ::|
Db  287  QKECAVFQMPDKDKESRICKALFSYMSHLRIYNDALIINEHARMKDALDYLRDFFRNIRA  346

Qy  636  KKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFPENNKMLKRLAENPE  695
              :::                                         ::||
Db  347  AGP----------------------DEIEQDL-----TQR----FEEKLQELESISIDPS  375

Qy  696  YENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLI  755
              :|||| || |||:| ||||||||||||:|||||||||||::||:|:|: :|
Db  376  NENPKLRDLCFILQEYHLNPET-RTILFVKTRALVDALKKWIKENPKLS--FLKPSILT  432

Qy  756  GAGHSSEFKPMTQNEQKEVISKFRTGKIN-LLIATTVAEEGLDIKECNIVIRYGLVTNEI  814
              ||| :|  |  |:|::::|:| :|| ||:  |||||||||||||| || :|||||||:|
Db  433  GRGKTNQNIGMTLPAQKCVLDTFRTDKDKILITTSVADEGIDIAQCNLVILYEYVGNVI  492

Qy  815  AMVQARGRARADES-TYVLVAH  835
              :||||| |   ||:  ||||||
Db  493  KMIQTRGRGRARGSKCFLLTAN  514
```

```
RESULT 11
US-09-864-761-45372
; Sequence 45372, Application US/09864761
; Patent No. US2002004876 3A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 45372
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010876.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.58
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.5
; OTHER INFORMATION: EST HUMAN HIT: AI906351.1, EVALUE 1.00e-42
; OTHER INFORMATION: SWISSPROT HIT: P14294, EVALUE 2.20e+00
US-09-864-761-45372

Query Match   14.7%;   Score 783; DB 9; Length 166;
Best Local Similarity   93.3%; Pred. No. 3.3e-49;
Matches   152; Conservative   3; Mismatches   8; Indels   0; Gaps   0;

```
Qy    1  MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE   60

Qy   61  LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL  120

Qy  121  LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESG  163
         ||||||||||||||||||:   :::
Db  121  LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRVGVCSDGASLLG  163
```

RESULT 12
US-10-106-698-5088
; Sequence 5088, Application US/10106698
; Publication No. US2003010690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5088
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (23)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5088

Query Match   14.7%;   Score 782; DB 14; Length 154;
Best Local Similarity   94.8%; Pred. No. 3.5e-49;
Matches   145; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

```
Qy  873  ILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMT  932
         ||||||||| |||| ||||  ||||||||||||||||||||||||||||||||||||||||
Db    2  ILELQMQSIXEKKMKXXXRNIAXHXXXNXPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMT   61

Qy  933  PEFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMMVHKGLDLPCLKIRNFVVVFK  992
         |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db   62  PEFKELYIVRENKXLQKKCADYQINGEIICKCGQAWGTMMVHKGLDLPCLKIRNFVVVFK  121

Qy  993  NNSTKKQYKKWVELPITFPNLDYSECCLFSDED  1025
         ||||||||||||||||||||||||||||||||
Db  122  NNSTKKQYKKWVELPITFPNLDYSECCLFSDED   154
```

RESULT 13
US-10-264-049-2244
; Sequence 2244, Application US/10264049
; Publication No. US2004000579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA13P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2244
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids

```
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (17)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (18)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (23)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (25)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (26)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (28)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
;    NAME/KEY: MISC_FEATURE
;    LOCATION: (75)
;    OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2244

Query Match     14.7%;  Score 782;  DB 15;  Length 154;
Best Local Similarity   94.8%;  Pred. No. 3.5e-49;
Matches  145;  Conservative  0;  Mismatches  0;  Indels  8;  Gaps  0;

Qy   873  ILELQMQSIMEKMKMKTKRNIAKHYKQNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMT 932
          |||||||||:||||:|||:|||:||:|||||||||||||||||||||||||||||||||||
Db     2  ILELQMQSIXEKXMKXXXRNIAXHXXNXPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMT 61

Qy   933  PEFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMMVHKGLDLPCLKIRNFVVVFK 992
          ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db    62  PEFKELYIVRENKXLQKKCADYQINGEIICKCGQAWGTMMVHKGLDLPCLKIRNFVVVFK 121

Qy   993  NNSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
          |||||||||||||||||||||||||||||||||
Db   122  NNSTKKQYKKWVELPITFPNLDYSECCLFSDED 154

RESULT 14
US-10-055-475-14
; Sequence 14, Application US/10055475
; Publication No. US2003022855A1
; GENERAL INFORMATION:
;   APPLICANT: Fisher, Paul B.
;   APPLICANT: Kang, Dong-Chul
;   APPLICANT: Gopalkrishnan, Rahul V.
;   TITLE OF INVENTION: USE OF MDA-5 AS AN ANTIVIRAL AND
;   TITLE OF INVENTION: ANTIPROLIFERATIVE AGENT
;   FILE REFERENCE: A34614-A-PCT-USA-A (070050.1921)
;   CURRENT APPLICATION NUMBER: US/10/055,475
;   CURRENT FILING DATE: 2002-01-22
;   PRIOR APPLICATION NUMBER: PCT/US01/06960
;   PRIOR FILING DATE: 2001-02-28
;   PRIOR APPLICATION NUMBER: 09/515,363
;   PRIOR FILING DATE: 2000-02-29
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-055-475-14

Query Match     11.6%;  Score 617;  DB 14;  Length 416;
```

```
Best Local Similarity   32.8%;  Pred. No. 1.8e-36;
Matches  174;  Conservative  77;  Mismatches  148;  Indels  132;  Gaps  15;

Qy   309  RPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIVLVNKVLL 368
          |:|:|||||:|||:||||:|||||||:.:::||| |:..:::|:..: ||: :..:::
Db     9  RNYQLELALPAMKGKNTIICAPTGCFKTFVSLLICEHHL--KKFPQGQKGKVVFFANQIPV 67

Qy   369  VEQ---LFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVKSCDIISTAQILENSLLNLE 425
          ||    .:: :|| :  ||  ...::|.|| |:...|.| :: .:.|: ||..|::|
Db    68  YEQQKSVFSKYFE---RHGYRVTGISGATAENVPEQIVENNDIIITPQIIVN---NLK 121

Qy   426  NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNRLKKENKPVIPLPQIL 485
          ||| .:: |:|.:|||||||  :::|:.|  |  |:|.: .::| |:::|| :||||
Db   122  KGTIP--SLSIFTLMIFDECHNTSKQHPYNMIMFNYLDQKLGGSS-------GPLPQVI 171

Qy   486  GLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT 545
          |||| |.||.|.|:| :: |:|:|||||:|||:|||.: : ..:.:   |  ||| | |
Db   172  GLTASVGVGDAKNTDEALDYICKLCASV----IATVKHNLELEQVVYKPQKFFRKVESR 227

Qy   546  REDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLR 605
          |||| |:.::     :|.::.:  .| :.:   :   :::.:
Db   228  ISDKFKYIIAQLMRDTESLA-----------------KRICKD--- 253

Qy   606  KYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDEYCDGDEDEDDLKK 665
               :::|:||:|:
Db   254  ----------------LENLSQIQNRE-------------------- 264

Qy   666  PLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEBSARGIIFT 725
               ::...                 :   : :   :||
Db   265  --KLQE---------------LESVSRDPSNENPKLEDLCFILQEBYHLNPETIT-ILFV 306

Qy   726  KTRQSAYALSQWITENEK--FAEVGVKAHHILGAGHSSEFKPMTONEQKEVISKFR-TGK 782
          ||.:   .|:|| :||:|     .:|:. ||:| .   ::|:.|.|: :.|: |: ||
Db   307  KTRALVDALKNWIEGNPKLSFLKPGI---LTGRGKTNQNTGMTLPAQKCIIDAFKASGD 362

Qy   783  INLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLV 833
          :|||||:|:.||||:|:|:|:||||||:.|:|:|:|||:|||: |:.|.
Db   363  HNILIATSVADEGIDIAQCNLVILYEYVGNVIKMIQTRGRGRARGSKCFLL 413

RESULT 15
US-10-264-049-2303
; Sequence 2303, Application US/10264049
; Publication No. US2004000557A1
; GENERAL INFORMATION:
;   APPLICANT: Birse et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PA133P1
;   CURRENT APPLICATION NUMBER: US/10/264,049
;   CURRENT FILING DATE: 2002-10-04
;   PRIOR APPLICATION NUMBER: PCT/US01/18569
;   PRIOR FILING DATE: 2001-06-07
;   PRIOR APPLICATION NUMBER: US 60/209,467
;   PRIOR FILING DATE: 2000-06-07
;   NUMBER OF SEQ ID NOS: 4360
;   SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2303
;   LENGTH: 447
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-264-049-2303

Query Match     10.6%;  Score 563.5;  DB 15;  Length 447;
Best Local Similarity   33.0%;  Pred. No. 1.6e-32;
Matches  155;  Conservative  75;  Mismatches  170;  Indels  69;  Gaps  16;

Qy   553  KLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKA---AKKGNRKERVC-----AE 602
          |:|| ||||||||||||||||||||||||||| :::||   ||||||||||
Db    13  KDLENLSQIQN------REFGTQKYEQWIVTVQKACMVFQMPDKDEESRICKALFLYTS 65

Qy   603  HLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDEYCDGDEDEDD 662
          |||| ::.|:||:|:|:|||:|:|||||||||:|||||||||||||||||||||||||
Db    66  HLRKYNDALIISEHARMKDALDYLKDFFSNVRAAGP----------------DEIEQD 107
```

```
Qy    663 LKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGI 722
               |       |       |   |  :   :  ::  :  |  |    ::  |:
Db    108 L-------TQR-----FEEKLQELESVSRDPSNENPKLEDLCFILQEEYHLNPETIT-I 153

Qy    723 IFTKTRQSAYALSQWITENEK--FAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFR- 779
               :|||||       ||   | : |:  |    |    |  :  :  |:
Db    154 LFVKTRALVDALKNWIEGNPKLSFLKPGI----LTGRGKTNQNTGMTLPAQKCILDAFKA 209

Qy    780 TGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSG 839
               : |:||||:|:|:||:||| |:|:||   |   || |:  |||||     :  |:|
Db    210 SGDHNILIATSVADEGIDIAQCNLVILYEYVGNVIKMIQTRGRGRARGSKCFLLT-SNAG 268

Qy    840 VIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHY--K 897
               |||   :  :||||       :    |   : | :|  |   :  : :
Db    269 VIEKEQINMYKEKQMNDSILRLQTWDEAVFREKILHIQTH---EKFIRDSQBKPKPVPDK 325

Qy    898 NNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQIN 957
               |:    |||      ||:  ||::    ||::       | :  :  :
Db    326 ENKKL---LCRKCKALACYTADVRVIEECHYTVLGDAFKECFVSRPHPK-PKQFSSFEKR 381

Qy    958 GEIIC---KCGQAWGTMMVHKGLDLPCLKIRNFVVFKNNSTKKQYKKW 1003
               :|:     |   |   | :  :|   ::| |||   :| ||
Db    382 AKIFCARQNCSHDWGIHVKYKTFEIPVIKIESFVVEDIATGVQTLYSKW 430
```

Search completed: June 2, 2004, 19:18:02
Job time : 57 secs

```
          GenCore version 5.1.6
     Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   June  2, 2004, 19:07:33 ; Search time 27 Seconds
                                     (without alignments)
                                     3651.715 Million cell updates/sec

Title:          US-09-515-363C-2
Perfect score:  5311
Sequence:       1 MSNGYSTDENFRYLISCFRA........LPITFPNLDYSECCLFSDED 1025

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:     283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    PIR.78:*
        1:    Pir1:*
        2:    pir2:*
        3:    pir3:*
        4:    pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	687	12.9		1037	2	T32534	hypothetical prote
2	654	12.3		811	2	T30968	helicase homolog C
3	642.5	12.1		956	2	A88708	protein C01B10.1 [
4	516	9.7		398	2	T46312	hypothetical prote
5	499	9.4		1134	2	T20332	hypothetical prote
6	453	8.5		752	2	D75219	ATP-dependent RNA
7	441.5	8.3		650	2	D71203	probable ATP-depen
8	417	7.9		741	1	A69432	ATP-dependent RNA
9	402.5	7.6		1374	2	S62524	probable RNA helic
10	392.5	7.4		784	2	H64487	eIF-4A family prob
11	350	6.6		1822	2	S44849	K12H4.8 protein -
12	342	6.4		738	1	E69055	ATP-dependent RNA
13	313.5	5.9		784	2	D84386	ATP-dependent RNA
14	303.5	5.7		993	2	S48436	probable RNA helic
15	284	5.3		502	2	T31323	ATP-dependent RNA
16	278.5	5.2		839	2	T39190	probable ATP-depen
17	223	4.2		557	2	G69958	SNF2 helicase homo
18	218	4.1		1979	2	C71622	hypothetical prote
19	217.5	4.1		2663	1	S28261	centromere protein
20	214.5	4.0		1790	2	S67593	transport protein
21	209.5	3.9		449	2	C71156	probable helicase
22	205	3.9		453	2	B75006	DNA repair protein
23	205	3.9		967	2	D72308	conserved hypothet
24	204.5	3.9		2163	2	S50675	pre-mRNA splicing
25	204	3.8		1939	2	T18372	repeat organellar
26	201.5	3.8		1063	2	T38420	probable DNA helic
27	198.5	3.7		2253	2	T30336	nuclear/mitotic ap
28	197.5	3.7		1163	2	G97236	ATPase involved in
29	195	3.7		1531	2	T48946	hypothetical prote
30	195	3.7		3259	1	A56539	giantin - human
31	194.5	3.7		986	2	E90596	restriction-modifi
32	194.5	3.7		1256	2	T26101	hypothetical prote
33	194	3.7		1039	2	E72734	hypothetical prote
34	193.5	3.6		663	1	G89848	exinuclease ABC su
35	193.5	3.6		2017	2	A36014	myosin heavy chain
36	193.5	3.6		2057	2	S61477	myosin II heavy ch
37	193.5	3.6		3225	2	I52300	giantin - human
38	193	3.6		715	2	B75135	DNA helicase relat
39	193	3.6		986	2	H90565	restriction modifi
40	192.5	3.6		718	2	E71100	hypothetical prote
41	192.5	3.6		1300	2	I53799	CG1 protein - huma
42	192	3.6		986	2	S49394	HsdR1 protein - My
43	192	3.6		1356	2	S32763	kinectin 1 - human
44	191.5	3.6		943	2	F69543	ATP-dependent RNA
45	191	3.6		1922	2	T00637	hypothetical prote

ALIGNMENTS

```
RESULT 1
T32534
hypothetical protein F15B10.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32534
R;Wohldmann, P.; Murray, J.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid F15B10.
A;Reference number: Z21186
A;Accession: T32534
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Experimental source: strain Bristol N2; clone F15B10
C;Genetics:
A;Gene: CESP:F15B10.2
A;Map position: 4
A;Molecule type: DNA
A;Residues: 1-1037 <WOH>
A;Cross-references: EMBL:AF036696; PIDN:AAB88350.1; GSPDB:GN00022; CESP:F15B10.2
A;Experimental source: strain Bristol N2; clone F15B10

Query Match             12.9%;  Score 687;  DB 2;  Length 1037;
Best Local Similarity   25.0%;  Pred. No. 1.4e-27;
Matches  251;  Conservative 169;  Mismatches 408;  Indels 174;  Gaps 35;

Qy    101 DLPSSFENAHDE--YLQLLNLLQPTLVDKLLVRDVLDKCMEEE----------- 142
Db     63 DLKNSILSNADDERLYKDIMTYLQ-TYLPKCTVHKLLN-CSNREVKLSDFHYILDHFEGF 120

Qy    143 LLTIEDRNRIAAAENNGNESGVRELLKRIVQK---ENWFSAFLN--VLRQTG--NNELVQ 195
Db    121 LRFIEPKVVIAYLDSYPQYIDAVAVLRKEIERNEEDNQDSDFIKKLILRTVPLLGEQAVY 180

Qy    196 EL--TGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMEN---------N 243
Db    181 DIMYTISEKSSNNLDVE-----------AKQFIA----------KVLRLKNDGFLRFYQIIN 221

Qy    244 SSESFFADSSVVSESDTSLAEGSVSCLDESLGHNS---NMGSDSGTMGSDSDE--ENV 296
Db    222 ASRRQLNGRIYICPVHESATEMMVYLGTAALNTNRYRMINIRVDNIVQENSTPRLVIESV 281

Qy    297 AARASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVVIAKDHLDKKKKASEP 356
Db    282 RQRIHRQRQLCLRNYQEELCQVALQGKNTIVTAPTGSGKTVIAANIIKEHPESRSSEGKR 341

Qy    357 GKVIVLV-NKVLLVEQLFRKEFQPPLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQ 415
Db    342 FKALFMTPNSMILNQQ--AASISSYLDHVHTQIIQGSD--NVPTRNVIQSKDLIVATPQ 397

Qy    416 ILENSLLNLENGEDAGVQ-----LSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNR 470
Db    398 MIVNLCNEHRNSLDDESRLDQFFLSFTIIFFDECHNTVKNSPYSNIMREY--HYILKNMG 455
```

Db   109 GKRFKALFMTPNSMILNQQETRISS------YLTQTIQGSD--NVPTRNVIQSKDLIVA 159

Qy   413 TAQILENSLLNLENGEDAGVQ------LSDPSLIIIDECHTTNKEAVYNNIMRHYLMQK 465

Db   160 TPQMIVN-LCN-EHRDPLDDEYPPEQFLSTFTIIFPDECHNTVKNSPYSNVMREY--HY 215

Qy   466 LKNNRLKKENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENL 525

Db   216 LKNMGNMPEGH--SFPQIIGLTASLGTGDKKNCMQVRSYIAGLCANMDVKELSIVKDNL 272

Qy   526 DQLKNQIQEPCKKFAIADATREDP------FKEKLLEIMTRIQTYCQMSPMSDFGTQP- 577

Db   273 EELLDHNPFVTDQVSFCERSNDGPIEMFTKRLKQMMQEVEDLIRTTLKNEPTVKYEIPPT 332

Qy   578 ------YEQWAIQMEKKCAAKKGNRKERVCAEHLRKYNEALQNDTIRMIDAYTHLETFYN 631

Db   333 DKEHNRYENWISNQRNCVSLAGSRNKTLIIEVL------DVLKVQYMFSQMKIFWN 382

Qy   632 EEKDKKFAV--IEDDSDEGGDDEYCDGDEDEDDLKKPLKLD------ETDRFLMTLFF 681

Db   383 YFRKYRFGFKRIFQFQDCFYALSYNINFNPEVALKKYLEKELGPERIRNFTDN--MNLIW 440

Qy   682 ENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITEN 741

Db   441 DN--CHRELVGIGSAENPMIARTVQFILDQNEQTSD--FRAIIFVRTKKEADFLNYVL--N 495

Qy   742 EKFAEVGVKAHHLIG---AGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDI 798

Db   496 DRLHELGIKSDWMSGQKKSTASSADISASKQKMEKLKMFADGENQILVSTSVAEEGLDI 555

Qy   799 KECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEH--ETVNDFREKMMYK 856

Db   556 PECSLVIKYNATNETAHVQRRGRARARNSKCVLITNS---IALHVQESNNLAKENLMTE 612

Qy   857 AIHCVQNMKPE------EYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNC 910

Db   613 TISLIQNSPGEFRQCVDEBSNKVWPRIQRETTDKAQRIKEQI----NRNIVYKIVCMKC 667

Qy   911 SVLACSGEDIHVIEKMHHVNMTPEF------KELYIVRENKALQKKCADYQINGBIIC--K 963

Db   668 DTVLCTNKDIR-SKNTQYIVCNPGFWSLVRRIPLPLEQRASNK----FNSTGSIECLGER 722

Qy   964 CGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVEL------PTTFPNLD 1014

Db   723 CGSKLGQLIDVNTVNLPCLKVKSILLLIESTNERIILVKQWKNILDEHFTPTTLKQRD 779

RESULT 3
A88708
protein C01B10.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: A88708
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elec
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: A88708
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-956 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAC47923.1; PIDN:g2911252; GSPDB:GN00022; CESP:C01B10.
C;Genetics:
A;Gene: C01B10.1
A;Map position: 4

Query Match   12.1%;   Score 642.5;   DB 2;   Length 956;
Best Local Similarity 25.0%; Pred. No. 2.4e-25;
Matches 253; Conservative 179; Mismatches 387; Indels 191; Gaps 39;

Qy   90 LAARYMNPE------LTDLPSPSFENAHDEYLQL----LNLLQPTLVDK 128

Qy   471 LKKENKPVIPLPQILGLTASPGVGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKN 530

Db   456 NMPEGH---SLPQIIGLTASLGTGDKNDCLQVRNYIAGLCASMDVKDLSIVKDNLEELRG 512

Qy   531 QIQEPCKKFAIADATREDP---FKEKLLEIMTRIQ   565

Db   513 YSPIVPDKVLLCERSTDGPIGMFTNRITLMMQEVEGLIRTALRNEHIGIEQRRQIETTER 572

Qy   566 QMSPMSDFGTQP------YEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTIR 618

Db   573 DFRPDSSFLDPPADKEHAGYQNWVC------NQMNLVSGTSFRETGTRTIINEALD 622

Qy   619 MI-DAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM 677

Db   623 VLKECFCTLSYNINFHIPEVALNYLKDEM------EY----RTPNFTV 659

Qy   678 TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW 737

Db   660 NMIRIKERYHNQLVGTGSAENPMISKTVQYIVEQNLQRADS-RTIIFVRTRYEATILNKV 718

Qy   738 ITENEKFAEVGVKAHHLIGAGHSSEFK---PMTQNEQKEVISKFRTGKINLLIATTVAEE 794

Db   719 LNSNEELLMLGIKSEWMSGLNKSTASSADISASKQMEKLKMFADGEIRILVSTSVAEE 778

Qy   795 GLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMM 854

Db   779 GLDVPECSLVIKYNATNETAHVQRRGRGRALNSECVLITNS-IALRDQESNNRDKESLM 837

Qy   855 YKAIHCVQNMKPE------EYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCK 908

Db   838 SETISLIQNSPAEFRKCVDEESNKIWPRILREDTDKAQKIEEQI---NRNIVYKIICK 892

Qy   909 NCSVLACSGEDIHVIEKMHHVNMTPEF---KELYIVRENKALQKKCADYQINGBIICK- 963

Db   893 KCEAILCTSKDIR-SRNTQYLVCDPGFWSLVRKTRILTDEQQALIK---YNATGSINCRR 947

Qy   964 --CGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKW 1003

Db   948 ENCGLKLGQLIEVNTVDLPCLSALSTVLLVBGTDKRIIVKKW 989

RESULT 2
T30968
helicase homolog C01B10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C;Accession: T30968
R;Blanchard, M.; Bradshaw, H.
submitted to the EMBL Data Library, August 1999
A;Description: The sequence of C. elegans cosmid C01B10..
A;Reference number: Z20949
A;Accession: T30968
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-811 <BLA>
A;Cross-references: EMBL:U58757; PIDN:AAD47117.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Map position: IV
A;Note: C01B10.1

Query Match   12.3%;   Score 654;   DB 2;   Length 811;
Best Local Similarity 27.0%; Pred. No. 4.9e-26;
Matches 210; Conservative 147; Mismatches 318; Indels 102; Gaps 27;
A;Introns: 28/2; 58/1; 113/3; 135/2; 165/2; 208/1; 285/1; 320/1; 442/1; 510/1; 606/3; 71

Qy   294 ENVAARASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKA 353

Db   49 ESVRQRIHIQRQFCLRNYQEELCQVALQGKNTIVTAPTGSGKTVIAANIIKEHFESRSSE 108

Qy   354 SEPGKVTVLV-NKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIIS 412

C;Accession: T46312
R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23036
A;Accession: T46312
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-398 <AAA>
A;Cross-references: EMBL:AL137608
A;Experimental source: adult testis; clone DKFZp434J1111
C;Genetics:
A;Note: DKFZp434J111.1

Query Match   9.7%; Score 516; DB 2; Length 398;
Best Local Similarity   33.3%; Pred. No. 2.2e-19;
Matches 141; Conservative 68; Mismatches 157; Indels 58; Gaps 14;

Qy   594 NRKERVC--------AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDE 647
                 :   : : ||||||||||||||| ::|:
Db   2 DEESRICKALFLYTSHLRKYNDALIISEHARMKDALDYLKDFFSNVRAAGF---------- 52

Qy   648 GGDDEYCDGDEDEDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNT 707
                 ::: ||| ||:|:||:||||||:|
Db   53 -------------DEIEQDL------TQR----FEEKLQELESVSRDPSNENPKLEDLCFI 90

Qy   708 IMEQYTRTEESARGIIFTKTRQSAYALSQWITENEK--FAEVGVKAHHLIGAGHSSEFKP 765
                 ||||||||||||||||||||||||||||||||||| : ::||| :
Db   91 LQEEVHLNPETIT-ILFVKTRALVDALKNWIEGNPKLSFLKPGI---LTGRGKTNQNTG 145

Qy   766 MTQNEQKEVISKFR-TGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRAR 824
                 ||||||||||||| ||||||||||:|||||||||::||||:||| :||||:||||||
Db   146 MTLPAQKCIILDAFKASGDHNILIATSVADEGIDIAQCNLVILYEYVGNVIKMIQTRGRGR 205

Qy   825 ADESTYVLVAHSGSGVIEHTVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEK 884
                 |||:||||||||||||||||||||||||||||||:||:|| |||||||||||||||
Db   206 ARGSKCFLLT-SNAGVIEKEQINMYKEKVMNDSIRLQTWDEAVFREKILHIQTH---EK 261

Qy   885 KMKTKRNIAKHY--KNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVR 942
                 :|: |: :| |::|||||||||||||:|||||||||||||||||||||||||:||
Db   262 FIRDSQEKPKPVPDKENKKL---LCRKCKALACYTADVRVIEECHYTVLGDAFKECFVSR 318

Qy   943 ENKALQKKCADYQINGEIIC--KCGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQ 999
                 |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   319 PHPK-PKQPSFEKFRAKIFCARQNCSHDWGIHVYKTFEIPVIKIESFVVEDIATGVQTL 377

Qy   1000 YKKW 1003
                 ||||
Db   378 YSKW 381

RESULT 5
T20332
hypothetical protein D2005.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20332
R;Wilkinson, J.
submitted to the EMBL Data Library, September 1996
A;Reference number: Z19258
A;Accession: T20332
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1134 <WIL>
A;Cross-references: EMBL:Z79752; PIDN:CAB02082.1; GSPDB:GN00019; CESP:D2005.5
A;Experimental source: clone D2005
C;Genetics:
A;Gene: CESP:D2005.5
A;Map position: 1
A;Introns: 33/2; 91/2; 116/3; 144/1; 319/1; 361/1; 421/3; 482/2; 591/1; 684/3; 757/1; 95

Query Match   9.4%; Score 499; DB 2; Length 1134;
Best Local Similarity   23.5%; Pred. No. 7.3e-18;
Matches 193; Conservative 134; Mismatches 288; Indels 206; Gaps 29;

Db   21 LESLYRDFEKGECFSELLPLGKIDELKTHS-ENSQEFSKQLYQDLKKSNLSIADP----E 75
                 | : :|| || ||:|| ||| | ::::::::: | :
Qy   129 LLVRDVL-------DKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAF 181

Db   76 RLYKDIMAYLQSNFPKCILHDVLNYSGRN-VKLSEFRYILDHLEGFLRLIEPK------VV 129
                 ||||||||||||||| : |||: | |||:|||:||||||:||||||| ||
Qy   182 LNVLRQTGN-NELVQELTGSDCSESNAEIENLSQVDGPQVEEBQLLSTTVQPNLEKEVWGM 240

Db   130 LNYIDSYPQYSEAVKLL--KDEIDRHEEDNN-----DPEFIKKLJLRTVPLLGEQAVVDI 182
                 || ||: :||| :|| :::|::||::| || |:|: ::| ||:||:| |
Qy   241 ENNSSESSFADSSVVSESDTSLA-EGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAAR 299

Db   183 MYTIWEKAQMTSIIINIRVDSIAPENSASRL------VIESVRQR 221
                 | :||| ::| ||:| ::|||| ||:: |:|:|:|:
Qy   300 ASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKV 359

Db   222 IHLQRQFCLRNYQEELCQVALQGKNTIVTAPTGSGKTVIAANIIKEHFESR---SSEG- 276
                 || || | || ::|| |:| ||| : |||||||| :|||| |:|:| |||
Qy   360 IVILVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPBVVKSCDIISTAQILEN 419

Db   277 -------------------------KRFKTKTQIGSD--NVPTRNVIQSKDLIVATPQMIVN 311
                                          |||| : : | ||| :|||:::|| :|||| |
Qy   420 SLNLENGEDAGVQ------------LSDFSLIIIDECHTNKEAVYNNIMRHYLMQKLKNNRLK 472

Db   312 -LCN-EHRDPLDDEYPEQFFLSTFTIIFFDECHNTVKNSPYSNVMREY--HYLKNMGNM 367
                    | ||:||::|| : | :|::|:|||||| ||| |||::|:: |||||:|:|
Qy   473 KENKPVIPLPQILGLTASPVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQI 532

Db   368 PEGH--SFPQIIGLTASLGTGDKKNCMQVRSYIAGLCANMDVKELSIVKDNLEELLDHN 424
                 || |||:|||||| |||| | : | : ||: ||:| |:|||:|:|||| |:| |
Qy   533 QEPCKKFAIADATREDP-------FKEKLLEIMTRIQTYCQMSPMSDFGTQP--------Y 578

Db   425 PFVTDQVSFCERSNDGPIEMFTKRLKQMMQEVEDLIRTTLKNEPTVKYBIPPTDKEHRY 484
                 || :| ::|| |:| || || |:| ||::|||| |||::||| |||:||||||| ||
Qy   579 EQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKF 638

Db   485 ENWISNQRNCVSLAGSRNKTLIIEVL----------DVLKVQYMFSQMKIFWNYFRKYRP 534
                 ||| |::|| | | ||:|| ||:|: |||::| ||::|| |||:|||: :
Qy   639 AV--IEDDSDEGGDDEYCDGDEDEDLKKPLKLD-------ETDRFLMTLFFENNKMLK 688

Db   535 GFKRIFQFQDCFYALSYNINFNPEVALKKYLEKELGPERIRNFIDN--MNLIWDN--CHR 590
                 ||||||| || ||| :|:|| ||| ||| |||:||| |||| | |::||| ||:
Qy   689 RLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVG 748

Db   591 ELVGIGSAENPMIARTVQFILDQNEQTSD-FRAIIFVRTKKEADFLNYVI--NDRLHELG 647
                 ||| |:|:||| |::| ||:|::: | ::|:||: ||| |:: ::|| |:|
Qy   749 VKAHHLIG--AGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVI 805

Db   648 IKSDWMSQQKKSTASSADISASKQMEKLKMFADGENQILVSTSVAEEGLDIPECSLVI 707
                 |||| |:|||:|::|:|||:| :| |||| |: ::|: ||||||||||| ||| |||
Qy   806 RYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEH-ETVNDFREKMMYKAIHCVQN 863

Db   708 KXNYATNETAHVQRRGRARARNSKCVLITNS---IALHVQESNNLAKENLMTETISLLQN 764
                 | |||| ||| ||||| |:|| |::|::   :|::|:|:|||| ||:|::||| ||
Qy   864 MKPE------EYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG 917

Db   765 SPGEFRQCVDEESNKVWPRIQREDTDKAQRIKEQI----NRNIVYKIVCMKCDTVLCTN 819
                 ||| :|| :|:: | |:||:| ||: || || ||||| :| |||||| ||
Qy   918 EDIHVIEKMHHVNMTPEF---KELYIVRENKALQKKCADYQINGEIIC--KCGQAWGT 970

Db   820 KDIR-SKNTQYIVCNPGFWSLVRRIPLPEQRASNK----FNSTGSIECLGERCGSKLGQ 874
                 ||:| |::| |:| | ||| | ||:| |:|: | :||||| ||| ||:| |
Qy   971 MMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVEL------PITFPNLD 1014

Db   875 LIDVNTVNLPCLKVKSILLLIESTNERILVKQWNILDEHFTPTTLKQRD 924
                 ||| ||||||||||| :||: |||||| |||:||| |:|||| ||| ||

RESULT 4
T46312
hypothetical protein DKFZp434J1111.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

```
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0190
C;Superfamily: ATP-dependent RNA helicase eIF-4A
F;33-40/Region: nucleotide-binding motif A (P-loop) #status atypical
F;111-136/Region: nucleotide-binding motif B
F;135-138/Region: DEAD/H motif

Query Match           8.5%;   Score 453;  DB 2;  Length 752;
Best Local Similarity 25.3%;  Pred. No. 9.4e-16;
Matches 183;  Conservative 124;  Mismatches 259;  Indels 156;  Gaps 28;

Qy 306 LQLREPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLV-N 364
           ::::|| ||   :|  :|::  |||||:|||::  :    ::  :   ::
Db  11 LQPRLYQ-ELIYAKCKERNCLIVLPTGLGKTIIAMMIADYRLNK-----YGGKVLMLAPT 64

Qy 365 KVLLVE--QLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVKS--CDIIISTAQILEN 419
       | |::      :|  ::|||:|                 ||:|
Db  65 KPLVLQHAETFRKFFNLPPEK---IVALTGE---VSPSERVKAWNRAKVIIATPQTVEN 117

Qy 420 SLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVVNNIMRHYLMQKLKNNRLKKENKPVI 479
       .::    :|:| :::||::: | | |:|   :  .:: :|:|:  :
Db 118 DLLT-----GKINLEDVVLVIFDEAHRAVGNYAVVYIAREYLNQ-AKN---------- 159

Qy 480 PLPQIIGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQBPCKKF 539
         .::|||||||           :::|::| :|:|
Db 160 --PHVIGLTASPG-------SSQEKILEVLRNLGIEHIYRSENSPDVKPYVQG--IKF 207

Qy 540 AIADATREDPFKE--KLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNR-- 595
           :.    ::|  |||:  :|      ||:
Db 208 EWIKVELPGLYKEVRKLLRDMLR----DSLKPLAEAGL--IESASADLPKKEILRAGQII 261

Qy 596 -KERVCAEH-LRK---YNEALQINDTIRMIDAY---THLETFYNEEKDKKFAVIED 643
               :|||| |||
Db 262 NEETAKGNHDLRKLLLFHAMALKLHHAIELLETQGLSALRVYLKKLYEEAK------ 312

Qy 644 DSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTK 703
                       ::  | | :     ::|:::  ||::
Db 313 ------------TGSTRASKELFLDRRMKKALALLIQ----AKELGIDHPKMEV 350

Qy 704 LRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENERKFAEVGVKAHHLIGAGHSSEF 763
       |::::|:|:::||  :  ::||  :   :: |::|||::
Db 351 LKELVKEQLSK-KENSKIIVFTNYRETARKVVEELTKE------GIKAKRFVGQATKEND 403

Qy 764 KPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR- 822
        .|:||||| :|.:| ||::|:::|:||:||||||::|::::||||||   | ||:|||
Db 404 RGMSQREQKLILDSFARGEFNVLVATSVGEEGLDVPEVDLVVFYEPVPSAIRSVQRRGRT 463

Qy 823 ARADESTYVLVAHSGSGVIEHETVNDFEREKMMYKAIHCV-------QNMKPEEYAHKILE 875
        |  | |  ::. |:       :::::| : :::  :|:            :|||::::
Db 464 GRQKPGRVVILIAQGTRDEAYYWSSKCERQMRETIRMVSQAIKREKQLSLESYVKKDTE 523

Qy 876 LQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEF 935
       ::::  ::|::.|.||:.|:||||  :  :|.::    :|:|:|::
Db 524 KEMEREEKEERKGEREGVKVVVDSRELRS------EVVKRLKTLGVKIEV 567

Qy 936 KEL----YIVRENKALQKKCADYQING-------------BIICKCGQAWGTMM 972
       | |    ::: |:   |:| |
Db 568 KTLDVGDYIISDEVAIERKSANDFIQSIIDGRLFDQVKRLKDSYPRPVVIVEGQLYGIRN 627

Qy 973 VH 974
       ||
Db 628 VH 629
```

```
Qy 294 ENVAARASPEP---------ELQLREPYQMEVAQPALEGKNIIICLPTGSGKTRVAVY 341
         : :   ::     :| :|     | | :||||||| :::  ||||||||:|||:|
Db 345 EDVCTTIRPQPKDSGYNPDAVVTEIVLRTIYQEELVQPALEGKNCVIVAPTGSGKTEVAIY 404

Qy 342 IAKDHLDKKKKASEPGKVIVLVNKVLLVEQLFRKEFQPLKKWYRVIGLSGDTQLKISFP 401
       :|| |:::   ::  |:::|| :: :: :|  : :   ::: |:
Db 405 AALKHIEERTSQGKPSRVLLVPKIPLVGQQ-KDRFLKYCNGMYEVNGPHG-SESSVSGT 462

Qy 402 --EVVKSCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIM 458
          :::  |:::      :| :: ::    |: |||| :: |:|:|
Db 463 GRRDEVIATHVSVMTPQILINMLQSVRQNE--RLYVSDFSMMIFDEVHKAAKNPY--VL 518

Qy 459 RHYLMQKLKNNRLKKENKPVIPLPQILGLTASPG--VGGATKQAKAEEHILKLCANLDAF 516
        ::::|::|::        .:.:.||||||:     ||:
Db 519 INQMVQEWKYEK----------PQIIGLTASLSVKVDGQKDENQMLNDIYNMLALINAP 567

Qy 517 TIKTV--KENLDQLKNQIQEPCKKFAIADATREDPFKE----------KLLEIM---- 558
        : :   :|:: :      .  :  :.            |||::
Db 568 HLSTITRQSSIDELNEHVGKPDDSVELCLPAKENILRDYIERYLNHAHGKFLEELASMSK 627

Qy 559 ----------TRIQTYCQMSPMSDFGTQPYE------QWAIQMEKKAAKKGNRKERVCAE 602
              ||: :      :    :        ::|:|::::
Db 628 STGRRNTIPPNMINTFKKKNQPKN------YEYYDSLLQGIIQELNKLNVPEKWNSQTWAK 681

Qy 603 HLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDD 662
            ::|| :  ::.:::  :::|  ::|
Db 682 YMKVYLEARGIVDLMPAMVAFKYME-------KAIGKLNEHSE--TVEYSTFIKDHDT 731

Qy 663 LKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGI 722
       ||: :    :  |:     :
Db 732 LKQTIQSVPE------------IVLRL------KKYT--HQSVPHQFGNYGEQMVGY 769

Qy 723 IFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGK 782
       :  : :: :   :: :
Db 770 VLGINKQGAV------------------QQTSQEQQLTLDKFNNGR 797

Qy 783 INLLIATTVAEEGLDIKECNIVIRYGLVT-NEIAMVQARGRARADESTYVLVAHSGSGVI 841
       :|::|:|:||||||::  |::::|:   | |  ::|||||||| ::| |  ::  |:| :
Db 798 LKVIVATSVVEEGLDVTACNLIIKINCSSGSAIQLVQQRGRARAKNSRSVLLSVK--SSIN 856

Qy 842 EHETVNDFEREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPS 901
       | ||  : :::   :  :  : :  ::| ::: ::::|::::|:
Db 857 ETETNALISEKYMRLCVKKITENGEKQLAAEVKRVAELNAAERKRNLEBQLNLRIRHENK 916

Qy 902 LITFLCKNCSVLACSGEDIHV--------IEKMHHVNMTPEFKELYIVRENKALQ 948
       ::  :|:||| :    ::     ::|::::  : |:
Db 917 IYKLMCSNCSKEFC--KSIYIKKVFSNYMVFDPSVWRFLHV----ESVETFI----KCLK 966

Qy 949 --KKC--ADYQIN-----------------GEIIC-KCG 965
         ::|  ||||:                  |:|: :|
Db 967 ITWKCRIADYQIAEFPNFAFQLTFRLFLCNFQMPQKRKVSKYLSEDNQPLSDIKCFHCK 1026

Qy 966 QAWGTMMVHKGLDLPCLKIRNFVVFKNNSTKKQYKKWVEL 1006
        | |         ::|:  :
Db 1027 LDVGRAYKIRGTYLFQLSVKALTFVQESDYSSMTKAKWSDV 1067
```

RESULT 6
D75219
ATP-dependent RNA helicase, eif-4a family PAB0190 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 10-May-2001
C;Accession: D75219
C;Status: preliminary
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: D75219
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-752 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49203.1; PID:g5457771

RESULT 7
D71203
probable ATP-dependent RNA helicase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 10-May-2001
C;Accession: D71203
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,

ATP-dependent RNA helicase homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-May-2001
C;Accession: A69432
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: A69432
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-741 <KLE>
A;Cross-references: GB:AE001002; GB:AE000782; NID:g2689325; PIDN:AAB89786.1; PID:g264910
C;Superfamily: ATP-dependent RNA helicase eIF-4A
C;Keywords: ATP; nucleotide binding; P-loop
F;36-43/Region: nucleotide-binding motif A (P-loop)
F;132-137/Region: nucleotide-binding motif B
F;136-139/Region: DEAD/H motif

Query Match   7.9%; Score 417; DB 1; Length 741;
Best Local Similarity 23.8%; Pred. No. 6.5e-14;
Matches 169; Conservative 114; Mismatches 230; Indels 198; Gaps 30;

```
Qy  303  EPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKKASEPGKVIVL  362
              :||::| ::: |   | :||:||:|||  :: |  | : : :  :
Db   11  ENTIERRMYQISIAATALT-KNTLVVIPTGLGKTTIALLVIASRL----LNEDGKVLFL   64

Qy  363  VNKVLLVEQLFRKEFQQPLKKWYRV---IGLSGDTQLKISFPEVVKSCDIIISTAQILEN  419
              |  |  :         ||||   ||:::    :::||:
Db   65  APTKPLIVEQHAR------FLKRVLKVBEIVSLSGEVPPE-KRKELWEKARIVVSTPQVVEN  118

Qy  420  SLLNLENGEDAG-VQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPV  478
              |||          | :::::::||  :  ||:: | | ::|
Db  119  DLL----------AGRISLEBDVILVVFDEAHRAVGNYAYVFIAKEYLRTAKK------  160

Qy  479  IPLPQILGLTASPGVGGATKQAKAEBEHILKLCANL--DAFTIKT------------  520
              |:| |||:||:|         || :|:::|:|  |
Db  161  ---PLIIAMTASPG---------SDPERIMEVIQNLGIEAIEVRTEWDSDVAPYVGKKRIE  209

Qy  521  --VKENLDQLKNQIQEPCKKFAIADATR----EDPFKEKLLEIMTRIQTYCQMSP  569
              :::| :| : ::| : | : :
Db  210  WIKVDIPEEMKEVKERLKE-CIKIRFKRLRELWIEVPENSSKRDLLALQEALQAEAASSQ  268

Qy  570  MSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDT-IRMIDAYTHLE  627
              |:|          : : :     :::::::||  :||| ::|::  | ||: | :| |
Db  269  SSEI----FEALSILAB------------IMKLQHAVELIETQGVKAVKSYLR--  305

Qy  628  TFYNEBKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKML  687
                       :|::||
Db  306  ------------------KLVRBATSKGG-----------------------SKAA  320

Qy  688  KRLAENP----------EYENEKLTKLRNTIMEQYTRTEBESARGIIFTKTRQSAYALS  735
              |:|:||           |:||:|||:|  | |:|:::|::::|||||:||:|:|:|:|
Db  321  KSIVGDPIFKKAVIALSKCKVEHPKLEKLKEILKEQEFEKNPDS-RVIVFTNYRDSAEML-  378

Qy  736  QWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEG  795
                   ||:|::|         : |:::| ||  | |  : ::|::|||:|||||:|:|||
Db  379  --VNELSPLFPVA------KFVGQASRDNDKGMRQKEQIETIDKPRRGVVKVLVATSVGEEG  432

Qy  796  LDIKECNIVIRYGLVTNEIAMVQARGR-ARADESTYVLVAHSGSGVIEHETVNDFREKMM  854
              |||: ::::::| |:  ||| ||:||| :|:|   : :|:  :|:::   | | :|:
Db  433  LDIPSTDLVVFYEAVPSEIRAIQRKGRTGRGREGRIVLVTKGT----------RDEAY  481

Qy  855  YKAIHCVQNMKPEYAH-KILELQMQSIMEKRMKTKRNIAKHYKNNPSLITFLCKNCSVL  913
              | :       |||:  :|  :    :  :| :::: : :  :::: ::   |:::::|
Db  482  Y-----YSSMKKERKMYDKILEI--KRIIDRK--QRSIGDYVLPEETGIKVIVDSRELR  531

Qy  914  ACSGEDIHVIEKMHHVNMTPEFKEL----YIVRENKALQKKCADYQINGEI  960
```

DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: D71203
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-650 <KAW>
A;Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31019.1; PID:g3258836
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
C;Gene: PH1896
C;Superfamily: ATP-dependent RNA helicase eIF-4A
C;Keywords: ATP; nucleotide binding; P-loop
F;33-40/Region: nucleotide-binding motif A (P-loop)
F;131-136/Region: nucleotide-binding motif B
F;135-138/Region: DEAD/H motif

Query Match   8.3%; Score 441.5; DB 2; Length 650;
Best Local Similarity 25.9%; Pred. No. 3e-15;
Matches 185; Conservative 122; Mismatches 261; Indels 147; Gaps 29;

```
Qy  306  LQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKKASEPGKVIVLV-N  364
              ||:|:||   : |:  :|| ::|||:|:||  | ::||:  |||          :|:| ::
Db   11  LQPRLYQ-ELIYAKKCKERNCLIVLPTGLGKTIIAMMIADYRLDK----YGGKVLMLAPT   64

Qy  365  KVLLVE-QLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKS---CDIIISTAQILEN  419
              | |:::   ||||||    |||::||  : ||:: |  ||:    | : ::| |:|||
Db   65  KPLVLQHAETFRKFFTLPPEK---IVALTGE---ISPNERVRAWARAKVVIATPQTIEN  117

Qy  420  SLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVI  479
              |||          ::|: |:::||: | ||:| | | : ::|:::|
Db  118  DLL------VGRISLEDVTLIIFDEAHRAVGNYAYVIAKEYLKQ-AKN------  159

Qy  480  PLPQILGLTASPGVGGATKQAKAEBHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKF  539
              | :::|||||:||      |   |:::::|:| ||| :::|
Db  160  --PHVIGLTASPG-------STPEKIMEVLRNLGIEHIEYRSENSPDVKPYVQG--IKF  207

Qy  540  AIADATREDPFKE--KLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNR--  595
              | :: :: :|::   |:|: : |:  : : :  ||::|||  | :|
Db  208  EWVKVELPELYKEVRKLLRDMLR-----DSLKPLAEAGL--LDSTSADIPKEILRAGQII  261

Qy  596  -KERVCAEH-LRK----YNEALQINDTIRMIDAY------THLETEYNEBKDKKFAVIED  643
                |::  ||  :|    |:|||  : : ::::::       | |:|:||:||
Db  262  NEEMAKGNHDLRKLLLFHAMALKLHAIELELTQGLSALRVYLKRLYEEAK-------  312

Qy  644  DSDEGGDDEYCDGDEDEDDLKKPLKLDETDRPLMTLFFENNKMLKRLAENPEYENEKLTK  703
              :::           ::::|:||:|| :  |  :::::| ::::||
Db  313  ------------AGSTRASKELFLDKRMKKALALLIQ-------AKELGIDHPKMEV  350

Qy  704  LRNTIMEQYTRTEBESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEF  763
              |:  :::|| | ::: |:|||||:| :|:|:|  : :| : : :||::::  :|: |:::
Db  351  LKGLIREQLKR-KENSKIIVFTNYRETAKKVV------EELMKDGIKARRFVGQASREND  403

Qy  764  KPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR-  822
              : |:::|||::: | ||::| :|:||:|::|||||:: ::::::| |:  ||| |||:||
Db  404  RGMSQKEQKLLDAFARSEFNVLVATSVGEEGLDVPEVDLVIFYEPVPSAIRSVQRRGRT  463

Qy  823  ARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIM  882
              :|:|:  : :|:  :|: :   |:::: |::||   : | ::|  |  ::: :| :::::
Db  464  GRQKPGRVILMAQGTRDEAYYWSSRQKEKIMRETIRMVSQM------VRKERQLSLESYV  518

Qy  883  EKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKEL----  938
              ::|: ::    :  ::|  |: : |  : :::   :|:::::|::::: : | | :|
Db  519  KREVNEE---IKEEKKETGGIKVVIDSRELRS-----EVKKLKTLGIKIEVRTLDVGD  569

Qy  939  YIVRENKALQKKCADYQING-----------------EIICKCCQAWGTMMVH  974
              |||:|::|::: :||:  |                     |:|| :|  |::|
Db  570  YIVSEDVAIERKSANDFIQSIIDGRLFDQVKRLKEAYPRPVIIEGGLYGIRNVH  624
```

RESULT 8
A69432

```
Db   532 S------EVVKGHLREIGAKIEIRNLEVADYVVSDRVAVERKTVEDFLNSII 576
```

RESULT 9
S62524
probable RNA helicase/ribonuclease SPAC8A4.08c - fission yeast (Schizosaccharomyces pomb
C;Species: Schizosaccharomyces pombe
C;Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 17-Nov-2000
C;Accession: T39130; T41432; T41192; S62524
R;Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data library, November 1995
A;Reference number: Z21830
A;Accession: T39130
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1374 <LYE>
A;Cross-references: EMBL:Z66569; NID:g1052533; PIDN:CAA91518.1; PID:g1052539; SPDB:SPAC8
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z21976
A;Accession: T41192
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 722-1374 <LYN>
A;Cross-references: EMBL:AL049662; PIDN:CAB41233.1; GSPDB:GN00068; SPDB:SPCC188.13c
C;Genetics: <LYE1>
A;Gene: SPAC8A4.08c
A;Map position: 1
C;Genetics: <SEE1>
A;Genetics: SPCC584.10c
A;Map position: 3
F;32-39/Region: nucleotide-binding motif A (P-loop) #status atypical
F;141-146/Region: nucleotide-binding motif B
F;145-149/Region: DEAD/H motif #status atypical

Query Match   7.6%; Score 402.5; DB 2; Length 1374;
Best Local Similarity 25.6%; Pred. No. 8.8e-13;
Matches 175; Conservative 107; Mismatches 244; Indels 157; Gaps 32;

```
Qy  304 PELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKK------KKA 353
Db    9 PQL-LRKYQQDVYNIASK-QNTLLVMRTGAGKTLLAVKLIIKQKLEEQILIQESNLEHKKI 66

Qy  354 SEPGKVIVLVNKVLLVEQ--LFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDII 410
Db   67 S-----VFLVNKVPLVFQQAEYIRSQLPAKVGMFYGELSIEMSEQL---LTNILKYNVI 118

Qy  411 ISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNR 470
Db  119 VITADLF---YLFLARG---FLSINDLNLIIFDECHHAIGNDAYARIMNDFY------HR 166

Qy  471 LKKE-NKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLK 529
Db  167 AKAVLSKKHFTLPRIFGMTASP-----------------------FTGK-KGNLYHRL 200

Qy  530 NQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKA 589
Db  201 YQ--------WEQLFDSKAHV--VSENELADYFCLPEESYVMYSNKLV 238

Qy  590 AKKGNRKERVCAEHL--------RKYNEALQINDTIRM----IDAYTHLETFYNEEKDKK 637
Db  239 VPPSDSIIKKCEETLQCKLISRAVKTALA--ETIDMGLWFGEQVWLYLVDFVETKRLKK 296

Qy  638 FAVIEDDSDEGGDEEYCDGDEDEDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYE 697
Db  297 KALGKQLSD------DEE----LAIDRLKIFVED--WKNNKYSDNGPRIPVFD 337

Qy  698 N----EKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHH 753
Db  338 STDVTDKVFKLLELLKATY-RKSDSVRTVIFVERKATAFTLSLFM--KTLNLPNIRAHS 393

Qy  754 LIGAGHS--SEFKPMTQNEQKEVISKPRTGKINLLIATTVAEEGLDIKECNIVIRYGLVT 811
Db  394 FIGHGPSDQGEFS-MTFRRQKDTLHKFKTGKYNVLIATAVAEEGIDVPSCNLVIRFNICR 452

Qy  812 NEIAMVQARGRARADESTYVLVAHSGSGVIE----HETVN---DFREKMYKAIHCV 861
Db  453 TVTQYVQSRGRARAMASKFLIFLNTEELLIHERILHEEKNLKFALSELSNSNIFDSLVCE 512

Qy  862 QNMK-PEEYAHKILE--------LQMQSIMEKKMKT-KRNIAKHY--KNNPSLITFLC- 907
Db  513 ERERVTDDIVEVGETGALLTGLYAVSLLYNFCNTLSRDVVTRYYPFTAQPCLSGWYCF 572

Qy  908 ------KNCSVLACSGEDIHVIEK 925
Db  573 EVELPKACKVPAAQGSPAKSIRK 595
```

RESULT 10
H64487
eIF-4A family probable ATP-dependent RNA helicase homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-May-2001
C;Accession: H64487
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, C
;Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: H64487
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-784 <BUL>
A;Cross-references: GB:U67591; GB:U67117; NID:g1592136; PID:g1592139; TIGR:MJ1505; PID:g1
C;Genetics:
A;Map position: REV1478605-1476251
A;Start codon: TTG
C;Superfamily: ATP-dependent RNA helicase eIF-4A
C;Keywords: ATP; nucleotide-binding binding; P-loop
F;41-48/Region: nucleotide-binding motif A (P-loop)
F;139-144/Region: nucleotide-binding motif B
F;143-146/Region: DEAH motif

Query Match   7.4%; Score 392.5; DB 2; Length 784;
Best Local Similarity 22.9%; Pred. No. 1.3e-12;
Matches 169; Conservative 128; Mismatches 235; Indels 205; Gaps 28;

```
Qy  306 LQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVVIAKDHLDKKKKASEPGKVIVLVNK 365
Db   19 LEARLYQQIIAANALKKKTLCV-LSTGLGKTAIALLVIAGLLTKK----DGKVLLIAPS 72

Qy  366 VLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
Db   73 RPLIVEQHYNRLKQVLNIDEDKIIALTGKIQPK-KRAELYKKGKIFIATPQVIENDII--- 128

Qy  426 NGEDAG-VQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQI 484
Db  129 ----AGRINVDBFILLIADEAHTTGHAY----AFVAKFPKDK------CHI 167

Qy  485 LGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLK----- 529
Db  168 LGLTASPG-----------SDIDKVMEICENLGIEHVBVRTEDDEDVKPYIAKVKLIPIRIDL 219

Qy  530 -NQIQEPCKKFAIADATRE------------DPFKEKLLEIMTRIQTYCQMSPMSDF 573
```

Db 220 PNEFKRAIK--LINEALKERLKILKDAGVINSIADVTKTELIELNNKLFSY------- 268
Qy 574 GTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTIRMIDA-------YTHLE 627
Db 269 --------------DEEVKYELIKVCSEALKLMHAKELLESQGKSVFLNYIN 306
Qy 628 TFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKML 687
Db 307 KLSMQRTKSAKSIVNDE----------------KVREAVNLLMKSDVEHPKL- 342
Qy 688 KRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEV 747
Db 343 -----GKVVDMVKNILEK-----NKDERIIIFAQYRDTVEKIVNLLIQN------- 381
Qy 748 GVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRY 807
Db 382 GIKAIRFIGQA-NKEGKGMSQKEQIEAIERFKK-EGSVLVSTSVSEEGIDIPSVNYIIFY 439
Qy 808 GLVTNEIAMVQARGRARADE--STYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNM- 864
Db 440 EPVPSEIRFIQRRGRAMRGEGGKVYVLLAKGTADEAYRSAL-YKEREMKRLL--KNMC 495
Qy 865 ----------KPEYAHKILELQMQSIMEKKMKTKRNIAKHYK------NNPSLITFLCKN 909
Db 496 YLLNKRLQKKFEKSKEEIKEETEEIKEKEIESKTAVKEETKEEEEKTKKPVTILDFIKQ 555
Qy 910 CSV--LACSGED------------IHVIEKMHHVNMTP------EFKEL---YIVRE 943
Db 556 IEVKERSKSEEDKIKQEIKIPKKPIKIIVDVREKNMAKLLHNYANIELKTLEVGDVYLSD 615
Qy 944 NKALQKKCADYQINGEI 960
Db 616 RVVVERKTAEDFVNSII 632

RESULT 11
S44849
K12H4.8 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
A;Reference number: S44851
C;Date: 14-Sep-1994 #sequence_revision 12-May-1994 #text_change 30-Jun-2001
A;Accession: S44849
R;Favello, A.D.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the C. elegans cosmid K12H4.
A;Molecule type: DNA
A;Status: preliminary
A;Residues: 1-1822 <FAV>
A;Cross-references: EMBL:L14331; NID:g289702; PIDN:AAA28101.1; PID:g289703
C;Genetics:
A;Introns: 17/3; 72/1; 116/3; 170/1; 216/3; 340/3; 401/3; 444/3; 512/2; 634/3; 694/3; 74
F;Keywords: ATP; nucleotide binding; P-loop
F;33-40/Region: nucleotide-binding motif A (P-loop)
F;141-146/Region: nucleotide-binding motif B
F;145-148/Region: DEXH motif

Query Match   6.6%;   Score 350;   DB 2;   Length 1822;
Best Local Similarity   23.5%;   Pred. No. 6.5e-10;
Matches   175;   Conservative   118;   Mismatches   249;   Indels   202;   Gaps   37;

Qy 309 RPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKDH----------LDKKKKASEPGKVIV 361
Db   14 RDYQVELLDKATK-KNTIVQLGTGSGKTFIAVLLKEYGVQLFAPLDQGGK-----RAFF 67
Qy 362 LVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDT-----QLKISFPEVVKSCDIIISTAQI 416
Db   68 VVEKVNLVEQ---QAIHIEVHTSFKVGGVHGQTSSGLWDSKEQCDQFMKRHHVVVITAQC 124
Qy 417 LENSLLNLENGEDAGVQLSDFSLIIIDECHHT-NKEAVYNNIMRHYLMQKLKNNRLKKEN 475
Db 125 L------LDLIRHAYLKIEDMCVLIFDECHHALGSQHPYRSIMVDY---------KLLKKD 170

Db 476 KPVIP.PQILGLTASPGVGVGATKQAKABEHILKLCANLDAFTIKTVKENLDQLKNQIQEP 535
Qy 171 KPV---PRVLGLTASL-IKAKVAPEKLMEQLKKLLESAMDS-VIETASD-LVSLSKYGAKP 224
Db 536 -----CKKFAIA--DATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQ- 576
Qy 225 YEVVIICKDFEIGCLGIPNFDTVIEIFDETVAFVNTTFEHPDLDLDPRRPIKDSLKTTR 284
Db 577 ----------PYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQIND-TIRMID-AYTHLE 627
Qy 285 AVFRQLGFPWAAW----------RTAQVWEKELGKIIKSQVLPDKTLRFLNMAKTSMI 331
Db 628 TFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDLKKPLKLDETDRFLMTLFFENNKML 687
Qy 332 TI----KRLL---------------EPEMKKIKSIEALRPYVPQRVLFEIL 365
Db 688 KRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYAL----SQWITE 740
Qy 366 ETF--NPEFQKERM-KLEKA----------EHLSAIIFVDQRYIAYSLLMMRHIKSW-- 410
Db 741 NEKFAEVGVKAHHLIGAG----HSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGL 796
Qy 411 EPKFK--FVNPDYVVGAGRNLASSDSQGL-HKRQTEVLRRFHRNEINCLIATSVLEEGV 467
Db 797 DIKECNIVIRYGLVTNEIAMVQARGRARADESTYLVAHSG----------SGVIE 842
Qy 468 DVKQCNLVIKFDRPLDMRSYVQSKGRARRAGSRYVITVEEKDTAAYCSKLPSDIFTRLVP 527
Db 843 HETVNDFREKMMYKAIHCVQNMKP--EEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
Qy 528 HNQIPIENGVTK--YCAELLLPINSPIKHAI---------VLKNPMPNKKT 569
Db 901 SLITFLCKNCSVLACSGE-DIHVIEK----------QKKCADY 954
Qy 570 AQMAVALEACRQLHLEGELDDNLLPKGRESIAKLLEHIDEEPDEYAPGIAAKVGSSKRKQ 629
Db 938 LYIVRENKAL----------QKKCADY 954
Db 630 LYDKKIARALNESFVEADKECFIY 653

RESULT 12
E69055
ATP-dependent RNA helicase, eIF-4A family - Methanobacterium thermoautotrophicum (strain
C;Species: Methanobacterium thermoautotrophicum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-May-2001
A;Accession: E69055
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
;Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: E69055
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-738 <MTH>
A;Cross-references: GB:AE000903; GB:AE000666; NID:g2622514; PIDN:AAB85892.1; PID:g262252
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1415
C;Superfamily: ATP-dependent RNA helicase eIF-4A
C;Keywords: ATP; nucleotide binding; P-loop
F;38-45/Region: nucleotide-binding motif A (P-loop)
F;134-139/Region: nucleotide-binding motif B
F;138-141/Region: DEXH motif

Query Match   6.4%;   Score 342;   DB 1;   Length 738;
Best Local Similarity   24.7%;   Pred. No. 4.6e-10;
Matches   172;   Conservative   106;   Mismatches   234;   Indels   184;   Gaps   31;

Qy 302 PEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIV 361

```
Db  12 PE-KIEARTYQQLLAADVIRKGNSMIVAPTALGKTVVAVLVAAERLRKYRGS----KVLI    66
           ::  :    ::     :     :::  ::||||  :: :        :
Qy 362 L-VNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISF-PEVVK----SCDIIISTAQ  415

Db  67 LSPSKPLAIQH--EESFREFM------LATCTSLTGSIKPEERKERWIKSQIISATPQ    116
         : : :  :     ::  ::  :    ||  :::: :|||  :::
Qy 416 ILENSLLNLENGEDAG-VQLSDFSLIIIDECHHTNKEAVNNIMRHYLMQKLKNNRLKKE  474

Db 117 TVESDIL------AGRYDLRDVSLIVFDECHRAVGSYVFLASNY---IQNAR-----    161
          :          |:  ::  :|||||| |: :   :         |:|
Qy 475 NKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANL-DAFTIKT----VKENLDQ   527

Db 162 ------HPLILGLTASPGAD----------EDKIKTVCENLFMNEVVVKTEGDPDVRPYLKP  207
          |::||||||||:       :: | |  : :: ||         :       :::|
Qy 528 LKNQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEK  587

Db 208 IKIEWVK----VRMTPELEDIRELLRKVLKN-RLKMLKNLGVID---TISVGK        252
         :         ::|   ::  ::  :  :   ::  ::|  |:|       |||
Qy 588 KAAKKG----NRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIE   642

Db 253 KDLLKARGRVQNRIARSTSPPRACYRAISLLASCINVEHALELLET-------------  298
          : :   |  ::|  ::: : : :: :: :::
Qy 643 DDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEY------  696

Db 299 ----QG----IRPL-----HQYLLRLKEKKTKAAKGLLADPDFTRAMHL            334
              |     |||          ::| | :::  :|||   | :: :::
Qy 697 -----ENEKLTKJRNTIMEQYRTEESARGIIFTKTRQSAYALSQWITENEKFAEV      747

Db 335 TRRAMMSGVEHPKLDRLME-ILKRELKGDE-ARIIVFTQFRDTLEEIYQ------RCKRE  386
            |    |    |:||::  ::|||:| :  |:::|||  : :||  :       | |:|
Qy 748 GVKAHHLIGACHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRY  807

Db 387 GINAVKFYGQNSRSGEKGLTQKQQRDIIKSFRMGNHDVLLSSVAEEGIDIPSVDLVVMY   446
           : | ::  |  :|  : | |    :: ::::|  | |:| |:||||| | |:|::|:
Qy 808 GLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPE  867

Db 447 EPVPSSIRMIQRRGRTGRKRKGRMVV---------LITEKTRD---EAYYYSSIRKERSMK-E  496
            |   :|:|:|:|:|||  |  |  |         ||:|| |    | :: :|||  :| |
Qy 868 EYAHKILELQMQSIMEKKMKYKRNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMH   927

Db 497 NLRGGSVNVEVNPIMEP-----------------                            525
          |       :| |:::|
Qy 928 HVN--MTPEFKELYIVRENKALQKKCADYQINGEII                          961

Db 526 EVNSRVIRELKKIGVDFELKPI--AVGDYQIISEDTI                         559
```

RESULT 13
D84386
ATP-dependent RNA helicase homolog eIF-4A [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 10-May-2001
C;Accession: D84386
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: D84386
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-784 <STO>
A;Cross-references: GB:AE004437; NID:g10581766; PIDN:AAG20456.1; GSPDB:GN00138
C;Genetics:
A;Gene: eif4a
C;Superfamily: ATP-dependent RNA helicase eIF-4A

Query Match 5.9%; Score 313.5; DB 2; Length 784;
Best Local Similarity 21.6%; Pred. No. 1.5e-08;
Matches 126; Conservative 91; Mismatches 190; Indels 175; Gaps 22;

```
Db 306 LQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK  365
           ::  ::   ::  :       |||||:||| :: :: : :          |::|:: |
Qy  18 IEARQYQLQLAAAARQG-HTLVCLPTGLGKTVSLLVTAYRL---ADDAGGTALLLAPT    72

Db 366 VLLVEQLFRKEFQPFLKKWYR------VIGLSGDTQLKISFPEVVKS----CDIIIS    412
          : :||| |::| ::                  |: :   : ::       |::
Qy  73 KPLIVEQ----------HAGFYREALALAIPDDDVVFTGETR------PDDRRAAWTDARVVVA  118

Db 413 TAQILENSLLNLENGEDAGVQ-SDFSLIIIDECHHTNKEAVVNIMRHYLMQKLKNNRLK   472
           | :: |  :  ::|      |  : :|||  |||  |  | ||  :  :    ||
Qy 119 TPQVVENDLVG------GRIDMDDVVHCTFDECHRATGDYAYTYVAERYHAD------   164

Db 473 KENKPVIPLPQILGLTASPGVGGATKQAKAEBHILKLCANLDAFTIKTVKENLDQLKNQI  532
                   | : | :|||         |     :| |  ||
Qy 165 ----------AAAPLVTAMSASPG----------GTEABIRTVCENLGVGNVEVMTED------  202

Db 533 QEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSMSDFGTQPYE--QWAIQMEKKA   589
            ::                                             : ::: |
Qy 203 ----------DA----------------DVGEHTHDTDVQW                       217

Db 590 AKKGNRKERVCAEHLRKYNEALQINDTIRMI-DAYTHLETF--------------YNEE  633
                         |         ::  :  :  |::
Qy 218 -------ERVTLP--------EEILEVRDAINDVIEDRLAKLREIGVTKASSPDISQKDLNEI  265

Db 634 KDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRF-LMTLFFEN--------  683
          : :         |  |   : |     :: |   :  :|     : : :
Qy 266 RARLQQLIDDDDSDGYQGMSVHAEVMK--LKRAVELVETQSVESVRRYFERQRNAANTAG  323

Db 684 -NKMLKRLLAEBNPEYENEKLT----------KLRNT--IMEQYRTEESARGIIFTKTRQSA  731
          :   :|   :  |         |:|      : :|    ::   : | ||
Qy 324 ASKASQRLVSEPAVKRAMRTAREFDGLHPKFRQARMLLAETLGIEDGRVIVFTESRDTA   383

Db 732 YALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTV  791
                 :|  :                          |  |   || |   | |:: |
Qy 384 EALTAFLGEH----------FDTRRFVGQGDADGSDGMTQTEQRETLAEFRNGDFEVLVSTSV  436

Db 792 AEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLV                   833
         |||||  |:     :  :  :|  :::|:|:|  : | :|
Qy 437 AEEGLDVPEVDLVLFFEPVPTAIRSVQRKGRGRQTAGRVVV                    478
```

RESULT 14
S48436
probable RNA helicase YIR002c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YIB2c
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
C;Accession: S48436; S50885
R;Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, August 1994
A;Reference number: S48432
A;Accession: S48436
A;Molecule type: DNA
A;Cross-references: GB:Z47047; EMBL:Z38062; NID:g9603997; PID:g763347; MIPS:YIR002c
A;Residues: 1-993 <BAD>
R;Voss, H.; Tamames, J.; Teodoru, C.; Valencia, A.; Sensen, C.; Wiemann, S.; Schwager, C
Yeast 11, 61-78, 1995
A;Title: Nucleotide sequence and analysis of the centromeric region of yeast chromosome I
A;Reference number: S50795; MUID:95282515; PMID:7762303
A;Accession: S50885
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-993 <VOS>
A;Cross-references: EMBL:X79743
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
C;Genetics:
A;Gene: SGD:MPH1
A;Cross-references: SGD:S0001441
A;Map position: 9R
C;Keywords: ATP; nucleotide binding; P-loop
F;107-114/Region: nucleotide-binding motif A (P-loop)
F;205-210/Region: nucleotide-binding motif B
F;209-212/Region: DEAD/H motif

```
Query Match          5.7%;   Score 303.5;  DB 2;  Length 993;
Best Local Similarity  20.3%;  Pred. No. 6.7e-08;
Matches  143;  Conservative  118;  Mismatches  246;  Indels  197;  Gaps   25;

Qy   307  QLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIV 366          Qy   303  EP-ELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIV 361
          :::||| ||  :|: ||||::|||||||||||:|:: :::  ::|                             |||                 :|: :|:   ::::  ::||
Db    86  EVRDYQYITVHKSL-FQNTLCAIPTGMGKTFIASTVMLNYFRWTKKA---KIIFTAPTR 140          Db    12  EPGAVERRDYQVGLAEQAIR-ENCIVVLPTGLGKTAVALQVIAHYLDEGRGALFLAPTRV  70

Qy   367  LLVEQLFRKEFQPFLKKWYRVIGLSGD------TQLKISFPEVVKSCDIIISTAQILENS 420          Qy   362  LVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQL--KISFPEVVKSC--DIIISTAQIL 417
          ||  |:     | : |:                    :::|  | :|                          ||||
Db   141  PLVAQQ------IKACLGITGIPSDQTAILLDKSRKNREEIWANKRVFFATPQVVEND 192          Db    71  LVN----QHRQFLG--RALTISDITILVTGEDTIPRRKAWGGSVICATPEIA 116

Qy   421  LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP 480          Qy   418  ENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKP 477
          |::               : :|:::                                                |::
Db   193  L------KRGVLDPKDIVCLVIDEAHRATGSSAYTNVVKF---------IDRFNSSY-- 234          Db   117  RNDI-----ERGLVPLEQFGLVIFDEAHRAVGDYAYSSIAR-----------------A 153

Qy   481  LPQIGLTASPGVGGATKQAKAEEHILKLCANLDAFTIK-TVKENLDQLKNQIQEPCKKF 539          Qy   478  VIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCK 537
          || | :||:                                                                : :|
Db   235  --RLLALTATP---------ASDLEGVQEVVNNLDISKIEIRTEESMDIVKYMKKRKKEKI 284          Db   154  VGDNSRMVGWTATL---PSEREKADE----IMGTLLSRSIAQRTEDDPDVKPYVQETAT 205

Qy   540  AIADATREDPFFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQM-------EKKAA 590          Qy   538  KFAIADATREDPFFKEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRK 596
          : :                                                                      :: :::
Db   285  EV------PLLLEIEDIIE------QLGMAVKPVLQQAIELGIYEECDPSQINAF 327          Db   206  EWIKVDLPPEMKEIQRLLKLALD-ERYSSLKRCGYDLGSNRSLSALLRL-RMVVLGGNRR 263

Qy   591  EHLRKYNEAL------QINDTIRMIDAYTHLETFYNEEKDKKFAV 640          Qy   597  ERVCAEHLRKYNEALQINDTIRMIDAY--THLETFYNEEKDKKFAVIEDDSDEGGDDEYC 654
          ::                                                                       : :
Db   328  KKGNRKERVCA---------EHLRKYNEAL------QINDTIRMIDAYTHLETFYNEEKDKKFAV 640 ... 
```

(Note: the above reproduces the densely printed alignment. Transcription of every residue in this two-column sequence alignment cannot be guaranteed letter-perfect.)

```
Qy   591  KKGNRKERVCA-------EHLRKYNEAL------QINDTIRMIDAYTHLETFYNEEKDKKFAV 640          Qy   597  ...
Db   328  KAMQQSQKIIANPTIPEGIKWRNFFILQLLNNVGQMLKRLKIY-GIRTFFNYFQNK---- 382          Db   264  A------AKPLFTAIRITYALNIFEAHGVTPFLKFCERTSKKKGVGVAE--------- 306

Qy   641  IEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFEN-NKMLKRLAEN-----PE 695          Qy   655  DGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTR 714
          |                    ||| ::|                                              ------
Db   383  --------CTEFTTKYNLKK-----STNKIAAEFYYHPILKNIKNQCENYLSDPK 424          Db   307  -------------LFEQDR----NFTGAIARAKAAQAGMEHPKIPKLEDAV-- 341

Qy   696  YENE-KLTKLRNTIMEQYTRTESARGIIFTKTRQSAYALSQWI------------- 738          Qy   715  TEESARG--IIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIG-AGHSSEFKPMTQNEQ 771
          |                                                                        ::::
Db   425  FVCHGKLQCVRDELMDFFQKRGSDSRVIIFTELRESALEIVKFIDSVADDQIRPHIFIGQ 484          Db   342  --RGARGKALVFTSYRDSVDLI------HSRLKAAGINSGILIGKAGE-----KGLKQRKQ 389

Qy   739  TENEKFAEVG-------VKAHHLIGAGH 759          Qy   772  KEVISKPRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR-ARADESTY 830
          :                                                                        :::
Db   485  ARAKEGFDEVKYTRKHAPKGRKKVERLHRQEQEKFLEAERTKRAANDKLERSARRTGSSE 544          Db   390  VETVAKFRDGGYDVLVSTRVGEEGLDISEVNLVIFYDNVPSSIRYVQRRGRTGRKDAGRL 449

Qy   760  SSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQA 819          Qy   831  VLVAHSGS 838
          |||                                                                      :::
Db   545  EAQISGMNQKMQKEVIHNFKKGEYNVLVCTSIGEEGLDIGEVDLIICYDTTSSPIKNIQR 604          Db   450  IVLMAKGT 457

Qy   820  RGR-ARADESTYVLVAHSGSVIEHETVNDFREKMMYKAIHCVQNMK---------PE- 867
          |
Db   605  MGRTGRKRDGKIVLLFSSNESYKFERAMEDYSTIQALISKQCIDYKKSDRIIPEDIIPEC 664

Qy   868  EYAHKILELQMQSIMEKKMKTKRNIAK 894
          :::
Db   665  HETLIITINDENEIINEMEDVDEVIRYATQCMMGKKKVKPKKAITK 708
```

Search completed: June 2, 2004, 19:12:27
Job time : 31 secs

RESULT 15
T31323
ATP-dependent RNA helicase, eIF-4A family - Cenarchaeum symbiosum
C;Species: Cenarchaeum symbiosum
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T31323
R;Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A;Title: Genomic analysis reveals chromosomal variation in natural populations of the un
A;Reference number: Z20994; MUID:98422450; PMID:9748430
A;Accession: T31323
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-502 <SCH>
A;Cross-references: EMBL:AF083072; NID:g3599393; PID:g3599409; PIDN:AAC62714.1

Query Match   5.3%;   Score 284; DB 2; Length 502;
Best Local Similarity 24.3%; Pred. No. 2.6e-07;
Matches 133; Conservative   94; Mismatches 207; Indels 114; Gaps   23;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2004, 19:01:12 ; Search time 18 Seconds
                                        (without alignments)
                                        2965.106 Million cell updates/sec

Title: US-09-515-363C-2
Perfect score: 5311
Sequence: 1 MSNGYSTDENFRYLISCFRA.........LPITFPNLDYSECLFSDED 1025

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	406	7.6	1909	1	DICE_ARATH	Q9sp32 arabidopsis
2	402.5	7.6	1374	1	YQ3D_SCHPO	Q09884 schizosacch
3	392.5	7.4	778	1	YF05_METJA	Q58900 methanococc
4	357	6.7	1912	1	DICE_HUMAN	Q9upy3 homo sapien
5	354	6.7	1906	1	DICE_MOUSE	Q8r418 mus musculu
6	350	6.6	1845	1	DCR1_CAEEL	P34529 caenorhabdi
7	303.5	5.7	993	1	YIS2_YEAST	P40562 saccharomyc
8	223	4.2	557	1	YQHH_BACSU	P54509 bacillus su
9	217.5	4.1	2663	1	CENE_HUMAN	Q02224 homo sapien
10	214.5	4.0	1790	1	USO1_YEAST	P25386 saccharomyc
11	205.5	3.9	1583	1	GCC2_HUMAN	Q8iwj2 homo sapien
12	204.5	3.9	2163	1	BRR2_YEAST	P32639 saccharomyc
13	197.5	3.7	1163	1	SBCC_CLOAB	Q97fk1 clostridium
14	197	3.7	1679	1	GCC2_MOUSE	Q8chg3 mus musculu
15	197	3.7	2230	1	GOA4_HUMAN	Q13439 homo sapien
16	195	3.7	3259	1	GOBL_HUMAN	Q14789 homo sapien
17	193.5	3.6	663	1	UVRB_STAAM	Q99v17 staphylococc
18	193.5	3.6	663	1	UVRB_STAAW	Q8nxm0 staphylococc
19	193.5	3.6	2017	1	MYSN_DROME	Q99323 drosophila
20	193	3.6	715	1	HELS_PYRAB	Q9v0a9 pyrococcus
21	192.5	3.6	1357	1	KTN1_HUMAN	Q86up2 homo sapien
22	191	3.6	3911	1	AKA9_HUMAN	Q99996 h a-kinase
23	189.5	3.6	3210	1	CENF_HUMAN	P49454 homo sapien
24	188.5	3.5	1433	1	REST_CHICK	042184 gallus gall
25	187.5	3.5	720	1	HELS_PYRFU	073946 pyrococcus
26	185	3.5	2611	1	BP1E_MOUSE	Q91zu8 mus musculu
27	183.5	3.5	663	1	UVRB_FUSNN	Q8rgr2 fusobacteri
28	183.5	3.5	1875	1	MLP1_YEAST	Q02455 saccharomyc
29	182.5	3.4	3660	1	DMD_CHICK	P11533 gallus gall
30	181.5	3.4	2869	1	RBP1_PLAVB	Q00798 plasmodium
31	181.5	3.4	479	1	DBPA_BACSU	P42305 bacillus su
32	180.5	3.4	976	1	SCP1_HUMAN	Q15431 homo sapien
33	180.5	3.4	1131	1	YAB9_YEAST	P31380 saccharomyc
34	180	3.4	796	1	DD27_HUMAN	Q96gq7 homo sapien
35	180	3.4	1690	1	C190_DROME	Q9vje5 drosophila
36	180	3.4	3214	1	BPA1_HUMAN	Q03001 homo sapien
37	179.5	3.4	607	1	DB10_NICSY	P46942 nicotiana s
38	179.5	3.4	1938	1	MYH6_MOUSE	Q02566 mus musculu
39	179.5	3.4	2022	1	ANT1_ONCVO	P21249 onchocerca
40	179	3.4	564	1	ROK1_YEAST	P45818 saccharomyc
41	179	3.4	662	1	UVRB_STRPN	Q54986 streptococc
42	179	3.4	662	1	UVRB_STRR6	Q8dpk7 streptococc
43	179	3.4	754	1	YAJ3_SCHPO	Q09903 schizosacch
44	179	3.4	2748	1	NUM1_YEAST	Q00402 saccharomyc
45	178.5	3.4	661	1	UVRB_STAEP	Q8cpz0 staphylococ

ALIGNMENTS

RESULT 1
DICE_ARATH
ID   DICE_ARATH   STANDARD;   PRT;   1909 AA.
AC   Q9SP32; Q9FDY6; Q9MANO;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Endoribonuclease Dicer homolog (EC 3.1.26.-) (CARPEL FACTORY protein)
DE   (SHORT INTEGUMENTS 1 protein) (SUSPENSOR1 protein).
GN   CAF OR SIN1 OR SUS1 OR AT1G01040 OR T25K16.4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND MUTANT CAF-1.
RC   STRAIN=cv. Wassilewskija;
RX   MEDLINE=20025923; PubMed=1055049;
RA   Jacobsen S.E., Running M.P., Meyerowitz E.M.;
RT   "Disruption of an RNA helicase/RNase III gene in Arabidopsis causes
RT   unregulated cell division in floral meristems.";
RL   Development 126:5231-5243(1999).
RN   [2]
RP   SEQUENCE FROM N.A., TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND
RP   MUTANTS SIN1-1 AND SIN1-2.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=22264007; PubMed=12376646;
RA   Golden T.A., Schauer S.E., Lang J.D., Pien S., Mushegian A.R.,
RA   Grossniklaus U., Meinke D.W., Ray A.;
RT   "SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY, a Dicer homolog, is a
RT   maternal effect gene required for embryo development in Arabidopsis.";
RL   Plant Physiol. 130:808-822(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu Q., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).

```
CC   -!- FUNCTION: Probably involved in cleaving double-stranded RNA in the
CC       RNA silencing pathway. It produces 20 to 23 bp dsRNAs (siRNAs and
CC       miRNAs) which target the selective destruction of homologous RNAs.
CC       Regulates by this way the early development of the embryo.
CC   -!- SUBUNIT: May interact with ARGONAUTE1 or PINHEAD through their
CC       common PAZ domains (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Highly expressed in flowers and seeds and
CC       detected in leaves and stems. Found in ovule integuments,
CC       inflorescence and floral meristems, stigma of flowers until just
CC       before pollination, vasculature of the funiculus, and embryo.
CC   -!- DEVELOPMENTAL STAGE: Detected in the embryo, but not in the
CC       suspensor, up to the globular stage.
CC   -!- MISCELLANEOUS: Expression in the early embryo is from the
CC       maternally contributed genome.
CC   -!- SIMILARITY: Belongs to the helicase family.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   -!- SIMILARITY: Contains 1 PAZ domain.
CC   -!- SIMILARITY: Contains 2 DRBM (double-stranded RNA-binding) domains.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF187317; AAF03534.1; -.
DR   EMBL; AF292941; AAG38020.1; -.
DR   EMBL; AF292940; AAG38019.1; -.
DR   EMBL; AC007323; AAF26461.1; ALT_SEQ.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR000999; RNase_III.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00035; dsrm; 2.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00358; DSRM; 2.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS50137; DS_RBD; 2.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   Hydrolase; Nuclease; Endonuclease; Helicase; Nuclear protein;
KW   ATP-binding; RNA-binding; Repeat.
FT   NP_BIND      249    256       ATP (POTENTIAL).
FT   DOMAIN      1180   1318       PAZ.
FT   DOMAIN      1342   1518       RNASE III 1.
FT   DOMAIN      1559   1707       RNASE III 2.
FT   DOMAIN      1733   1796       DRBM 1.
FT   DOMAIN      1831   1906       DRBM 2.
FT   SITE         378    381       DECH BOX.
FT   MUTAGEN      415    415       P->S: IN SIN1-1; IMPAIRED REPRODUCTIVE
FT                                 DEVELOPMENT.
FT   MUTAGEN      431    431       I->K: IN SIN1-2; IMPAIRED REPRODUCTIVE
FT                                 DEVELOPMENT.
FT   MUTAGEN     1837   1843       NDICLRK->IAEIDPG: IN CAF-1; CONVERTS
FT                                 THE FLORAL MERISTEM TO AN INDETERMINATE
FT                                 STATE.
FT   MUTAGEN     1844   1909       MISSING: IN CAF-1; CONVERTS THE FLORAL
FT                                 MERISTEM TO AN INDETERMINATE STATE.
FT   CONFLICT     148    148       S -> F (IN REF. 1 AND 2).
```

```
FT   CONFLICT     988    988       Y -> H (IN REF. 1 AND 2).
SQ   SEQUENCE   1909 AA;  213572 MW;  EAA944F0C0C81D4C CRC64;

   Query Match             7.6%;  Score 406;  DB 1;  Length 1909;
   Best Local Similarity   23.0%;  Pred. No. 2.2e-12;
   Matches  199;  Conservative  131;  Mismatches  271;  Indels  264;  Gaps  40;

Qy    135 LDKCME------BELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAFLNVLRQT 188
              ::| :        ::::  :  :      : : ::: :
Db     18 LDACEDISCDLIDDLVSEFDPSSVAVNEST-DNGV------INDFFGGIDHILDSI     67

Qy    189 GNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPN-LEKEVWGMENNSSES 247
              :||                      :    :           :
Db     68 KNG--GGLPNNGVSDTNSQINEVTVT-PQV--IAKETVKENGLQKNGGKRDEFSKEE   119

Qy    248 SFAD------SSVVSESDTSLABGSVSCLDE---------SLGHN------         277
              :|                                    |:
Db    120 GDKDRKRARVCSYQSERSNLSGRGHVNNSREGDRFMNRKRTNWDEAGNNKKKRECNNYR 179

Qy    278 ------------SNMGSD------SGTMGDSD------------EENVAARAS         301
                          :||||          |:
Db    180 RDGRDREVRGYWERDKVGSNELVYRSGTWEADHERDVKKVSGGNRECDVKAEEN---KSK 236

Qy    302 PEP------ELQLRPYQMEVAQPALEGKNIIICLPTGSGKIRVAVY----IAKDHLDKKK 351
              | :      : ::|   :  : :     : | ||: ||| :
Db    237 PEERKEKVVEBQARRYQLDVLEQA-KAKNTIAFLETGAGKTLIAILLIKSVHKDLMSQNR 295

Qy    352 KASEPGKVIVLVNKVLLVEQ----LFRKEF-------------QPF--LKKWYRVIGLSGDT 394
                         ::| :|  :|              :|:   |
Db    296 KMLS----VFLVPKVPLVYQQAEVIRNQTCFQVGHYCGEMGQDFWDSRRWQREF------  345

Qy    395 QLKISFPEVVKSCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVY 454
                          : :|||:   :::  : ::|                |||| ::
Db    346 ------ESKQVLVMTAQILLNIL------RHSIIRMETIDLLILDECHAVKKHPY        389

Qy    455 NNIMRHYLMQKLKNNRLKKENKPVIPLPQIIGLTASP-GVGGATKQAKABEHILKLCANL 513
              ::|:|::                    | : ||:|||  : :          :: |:
Db    390 SLVMSEFYHTTPKDKR---------PAIFGMTASPVNLKGVSSQVDCAIKIRNLETKL   438

Qy    514 DAFTIKTVKENLDQLKNQIQEPCKKFAIAD-ATREDPFKEKLLEIMTRIQTYCQMSPMSD 572
              | :  |:|          |     |::|       ::   ::::  | ::
Db    439 DS-TVCTIKDR-KELEKHVPMSEIVVEYDKAATMWSLHETIKQMIAAVEEAAQASSRK-  495

Qy    573 FGTQPYEBQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTIRMID------A    622
                          |: |  ||   :      :        :  :
Db    496 ------SKWQFMGARDAGAKDELRQVYGVSERTESDGAANLIHKLRAINYTLAELGQWCA 549

Qy    623 YTHLETF-----------YNEEKDKKF                                642
              : |:                :
Db    550 YKVGQSFLSALQSDERVNFQVDVKFQESYLSEVVSLLQCELLEGAAAEKVAAEVGKPENG 609

Qy    643 ---DDSDEG--GDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYE 697
                 ::: :|   |:                              : ::
Db    610 NAHDEMEEGELPDDPVVSGGEHVDEVIGAAVAD-----------GKVTPKV------   649

Qy    698 NEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWTENEKFAEVG---VKAHH   753
                : :|              : |:   ||  :                      |:|
Db    650 -QSLIKL----LLKYQHTAD-FRAIVFVERVVAALVLP------KVFAELPSLSFIRCAS 697

Qy    754 LIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKBCNIVIRYGLVTNE 813
              :||| : |:         ||   ||||| :|||||:|:||||||| ||::|:| | :||
Db    698 MIGHNNSQEMK-----SSQMQDTISKFRDGHVTLLVATSVAEEGLDIRQCNVVMRFDLAKTV 754

Qy    814 IAMVQARGRARADESTYVLVAHSGS                                  838
              :| :|:|||||  || |:|: : |:
Db    755 LAYIQSRGRARKPGSDYILMVERGN                                  779


RESULT 2
YQ3D_SCHPO
ID   YQ3D_SCHPO     STANDARD;      PRT;  1374 AA.
AC   Q09884; Q9UUN1;
DT   01-FEB-1996 (Rel. 33, Created)
```

```
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative helicase C18B.13c (EC 3.6.1.-).
GN   SPCC188.13C OR SPCC584.10C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SIMILARITY: Belongs to the helicase family.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AL033824; CAB37423.1; -.
DR   EMBL; AL049662; CAB41233.1; -.
DR   PIR; T39130; S62524.
DR   GeneDB_SPombe; SPCC188.13c; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR000999; RNase_III.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS00517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   Hypothetical protein; Helicase; ATP-binding; Hydrolase; Nuclease;
KW   Endonuclease; Repeat.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   SITE        145    148       DECH BOX.
FT   DOMAIN      916   1038       RNASE III 1.
FT   DOMAIN     1083   1233       RNASE III 2.
```

```
SQ   SEQUENCE   1374 AA;  158039 MW;  89AE9EF8DE7966C6 CRC64;

Query Match             7.6%;
Best Local Similarity  25.6%;   Score 402.5; DB 1;  Length 1374;
Matches  175;  Conservative 107;  Mismatches 244;  Indels 157;  Gaps  32;

Qy     304 PELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKK------------KKA 353
               ::||:|||:|: ||:  |||:|:|:|:|||||:|::|:  |:|:         |||
Db       9 PQL-LRKYQQDVVNIASK-QNTLLVMRTGAGKTLLAVKLIKQKLEEQILIQESNLEHKKI  66

Qy     354 SEPGKVIVLVNKVLLVEQ---LFRKEFQPFLKKWYRVIGLSGDTQLKISFPPEVVKSCDII 410
               ::|:|:||||  :        :: |:|||::|:|:||: |:|:|::||:||: :::|:|
Db      67 S-----VFLVNKVPLVFQQAEYIRSQLPAKVGMFYGELSIEMSEQL---LTNILLKYNVI 118

Qy     411 ISTAQILENSLLNLENGDAGVQLSDFSLIIIDECHHTNKEAVNNIMRHYLMQKLKNNR 470
               :: :|:|: :|| | :  ::::||:|||||||||:| :::  :|:||:::|:|:|:
Db     119 VITADLF--YLFLARG--FLSINDLNLIIFDECHHAIGNDAYARIMNDFY-----HR 166

Qy     471 LKKE-NKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLK 529
               |:|:  ||:: |:|: :|||:||| |:| : |: |||:: ::|:|:|::|||::|:|:
Db     167 AKAVLSKKHFTLPRIFGMTASP----------------FTGK--KGNLYHRL 200

Qy     530 NQIQEPCKKFAIADATREDPFKEKLLEIMTRIQYCQMSPMSDFGTQPYEQWAIQMEKKA 589
               :|:|     :|:::::: :: :::::|::||::|::::::::::||   :::::|::|
Db     201 YQ----------------------WEQLFDSKAHV--VSENELADYFCLPEESYVMYSNKLV 238

Qy     590 AKKGNRKERVCAEHL-----------------RKYNEALQINDTIRM--------IDAYTHLETFYNEEKDKK 637
               ||:|:|: |::|:|:                  ||::|:::|: |:         |:|||:|::|:||:|:|:
Db     239 VPPSDSIIKKCEETLQGCKLISRAVKTAlA--ETIDMGLWFGEQVWLYLVDFVETKRLKK 296

Qy     638 FAVIEDDSDEGGDDEYCDGDEDDLKKPLKLDETDRFLMTLFPENNKMLKRLAENPEYE 697
               ::|:|  |::|:   ::::: :::   ::|:|::|:|::|:|:::|:|::|::|:|::
Db     297 KALGKQLSD---------DEE------LAIDRLKIFVED--WKNNKYSDNGPRIPVFD 337

Qy     698 N----EKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHH 753
               |    |::|::|: ||::|:|:|::||||||:|:||::|||:::||:|:|:|:|:|:|:|
Db     338 STDVTDKVFKLLELLKATY-RKSDSVRTVIFVERKATAFTLSLFM--KTLNLPNIRAHS 393

Qy     754 LIGAGHS---SEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVT 811
               ||| |::    :|::|||:|::|:|:|||||:|:|||||:|||:|:|::|||:|:||:|:|
Db     394 FIGHGPSDQGEFS-MTFRRQKDTLHKFKTGKYNVLIATAVAEEGIDVPSCNLVIRFNICR 452

Qy     812 NEIAMVQARGRARADESTYVLVAHSGSGVIE----HETVN------DFRRKMMYKAIHCV 861
               ||::|:|:|| |:|::|:|:||:|:|:|::      |::||      |:|:|::|:|:|:|
Db     453 TVTQYVQSRGRARAMASKFLIFLNTBELLIHERILHEEKNLKFALSELSNSNIFDSLVCE 512

Qy     862 QNMK-PEEYAHKILE-------LQMQSIMEKKMKT-KRNIAKHY---KNNPSLITFLC- 907
               |:|:  |::|:|::         ::||:::|::||: |||:||:|    |:|||:|:||:
Db     513 ERERVTDDIVYEVGETGALLTGLYAVSLLYNFCNTLSRDVTRYYPTFTAQPCLSGWYCF 572

Qy     908 -----KNCSVLACSGEDIHVIEK 925
                    |||:|:|:|:::::|:::
Db     573 EVELPKACKVPAAQGSPAKSIRK 595
```

```
RESULT 3
YF05_METJA
ID   YF05_METJA      STANDARD;      PRT;    778 AA.
AC   Q58900;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative ATP-dependent RNA helicase MJ1505.
GN   MJ1505.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
```

```
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: Belongs to the DEAD box helicase family. DEAH
CC       subfamily.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U67591; AAB99518.1; -.
DR   TIGR; MJ1505; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_box.
DR   InterPro; IPR006166; ERCC4.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000445; HhH.
DR   InterPro; IPR003583; HhH_1.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF02732; ERCC4; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00633; HhH; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00278; HhH1; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   Hypothetical protein; ATP-binding; RNA-binding; Helicase;
KW   Complete proteome.
FT   NP_BIND      35     42       ATP (POTENTIAL).
FT   SITE        137    140       DEAH BOX.
FT   DOMAIN      532    535       POLY-GLU.
SQ   SEQUENCE   778 AA;  88882 MW;  DCB1F22080133D8D8 CRC64;

Query Match             7.4%;  Score 392.5;  DB 1;  Length 778;
Best Local Similarity  22.9%;  Pred. No. 3e-12;
Matches  169;  Conservative  128;  Mismatches  235;  Indels  205;  Gaps  28;

Qy    306 LQLRPYQMBVAQAPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIVLVNK 365
              ||| ::          :   || ||:||:||        ::  |      |::| ::|
Db     13 LEARLYQQIIAANALKKKTiCV-LSTGLGKTAIALLVIAGLLTKK-----DGKVLILAPS 66

Qy    366 VLLVEQLFRKEFQPPLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
              :||| :  ||    ||   :   | ||:    :    :     |||||  |  :||
Db     67 RPLVEQHYNRLKQVLNIDEDKIIALTGKIQPK-KRAELYKKGKIFIATPQVIENDII--- 122

Qy    426 NGEDAG-VQLSDFSLIIIDECHHTNKEAVYNIMRHYLMQKLKNNRLKKENKPVIPLPQI 484
                  |  : |  ::|||:| ||||  :   |  ::      | : |
Db    123 ---AGRINVDEFILLIADEAHHTGDHAY-----AFVAKKFKDK---------------CHI 161

Qy    485 LGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLK------------- 529
          ||||||||       || :  || ::|| | : :   :  :
Db    162 LGLTASPG------SDIDKVMEICENLGIEHVEVRTEDEDVKPYIAKVKLIPIRIDL 213

Qy    530 -NQIQEPCKKFAIADATRE----------DPFKEKILEIMTRIQTYCQMSPMSDF 573
            : : |  :|     ||  ::  |:  :| |  :| |
Db    214 PNEFKRALK-LINEALKERLKILKDAGVINSIADVTKTELIELNNKLFSY------- 262

Qy    574 GTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQNDTIRMIDA------YTHLE 627
          |  | ||    :|          : :|        |:     :
Db    263 ---------------DEEVKYELIKVCSSALKMHAKELLESQGKSVFLNYIN 300

Qy    628 TFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKML 687
                :   :|:              :   |         |      |:
Db    301 KLSMQRTKSAKSIVNDE-------------------KVREAVNLLMKSDVEHPKL- 336

Qy    688 KRLAENPEYENEKLTKLRNTIMEQYRTEESARGIIFTKTRQSAYALSQWITENEKPAEV 747
           :   ::       |  ::    |    ::|     |     :         ::
Db    337 ------GKVVDMVKNILEK---NKDERIIIFAQYRDTVEKIVNLLTQN----- 375

Qy    748 GVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRY 807
          |:|    ||    :    : | ||:| | :  :|||| :  ::  ||| | : |::|  :
Db    376 GIKAIRFIGQA-NKEGKGMSQKEQIEAIERFKK-EGSVLVSVSVSSEGIDIPSVNYIIFY 433

Qy    808 GLVTNELAMVQARGRARADE--STYYLVAHSGSGVIEHETVNDFRRKMMYKAIHCVQNM- 864
             :  ||  ||||| ||      | |: : |:|:| :  ::  || :|:||  :|::
Db    434 EPVPSEIRFIQRRGRAMRGEGKVYVLIAKGTADEAYYRSAL-YKEREMKRLL---KNMC 489

Qy    865 ------KPEEYAHKILELQMQSIMEKKMKTKRNIAKHYK------NNPSLITFICKN 909
               ||||  |: |:|:  :|||:|||||:|| |:|||        :  ||| :|:|
Db    490 YLLNKLRLQKKFEEKSKEEIKEETEEIKEKEIESKTAVKEETKEEEKTKKPVTILDFIKQ 549

Qy    910 CSV-LACSGED------------IHVIEKMHHVAMTP-------EFKEL---YIVRE 943
           :: |||| ||            ||:||:: | | ||       |::|      ::||
Db    550 IEVKERSKSEEDKIKQEIKIPKKPIKIIVDVREKNMAKLLHNYANIELKTLEVGDYVLSD 609

Qy    944 NKALQKKCADYQINGEI 960
             :| :| |:  :  ::
Db    610 RVVVERKTAEDFVNSII 626


RESULT 4
DICE_HUMAN
ID   DICE_HUMAN     STANDARD;      PRT;   1912 AA.
AC   Q9UPY3; O95943; Q9UQ02;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endoribonuclease Dicer (EC 3.1.26.-) (Helicase with RNase motif)
DE   (Helicase-MOI).
GN   DICER1 OR DICER OR HERNA OR KIAA0928.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20246304; PubMed=10786632;
RA   Matsuda S., Ichigotani Y., Okuda T., Irimura T., Nakatsugawa S.,
RA   Hamaguchi M.;
RT   "Molecular cloning and characterization of a novel human gene (HERNA)
RT   which encodes a putative RNA-helicase.";
RL   Biochim. Biophys. Acta 1490:163-169(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Provost P., Dishart D., Doucet D., Hermansson A., Frendewey D.,
RA   Samuelsson B., Radmark O.;
RT   "RNA binding and processing by recombinant human Dicer.";
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
RN   [4]
RP   REVISIONS.
RX   MEDLINE=22158633; PubMed=12168954;
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
```

```
Query Match                  6.7%;  Score 357;  DB 1;  Length 1912;
Best Local Similarity   24.6%;  Pred. No. 5.9e-10;
Matches  160;  Conservative  88;  Mismatches 203;  Indels 200;  Gaps  29;

Qy  309  RPYQMEVAQPALEGKNIICLPTGSGKTRVAVYIAKD-HLDKKKKASEPGK-VIVLVNKV  366
              :|::|:.|:|| ||||||||.|:|..|   |   | ||||      |:|| |:|.||||
Db   35  RKYQVELLEAALD-HNTIVCLNTGSGKTFIAVLLTKELSYQIRGDFSRNGKRTVFLVNSA   93

Qy  367  LLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKI---SFPEVVKSC------DIII      411
              :  :|||  ||  :::|::::   |  | :|     || ||  |:     ::|
Db   94  NQVAQ------------QVSAVRTHSDLKVGEYSNLEVNASWTKERWNQEFTKHQVLI    139

Qy  412  STAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRL  471
                  ||:..|:.||.|    :|||  ::||:||| |       |  :: ::|
Db  140  MTCYVALNVLKN------GYLSLSDINLLVFDECHLALLDHPYREIMK-            181

Qy  472  KKENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQ-    527
                ::|  :|||||||| : :||| |||::|:|:| :| |     :|:|::|
Db  182  LCENCP-SCPRILGLTASI-LNGKCDPEELEKIQKL------EKILKSNAETATDL    230

Qy  528  --LKNQIEPCKKFAIAD---ATREDFKEKLLEI-------MTRIQTYC           565
               : |:|::|  ||:     ::|:|:|:||::|        :|: :
Db  231  VVLDRYTSQPCE--IVVDCGPFTDRSGLYERLLMELEEALNFINDCNISVHSKERDSTLI  288

Qy  566  QMSPMSD-------FGTQPYEQWAIQMEKKAAK--KGNRKE------RV           599
              .:|:.||        |.:|  ||::|:::::     |||:|
Db  289  SKQILSDCRAVLVVLGPMCADKVAGMMVRELQKYIKHEQEELHRKFLLFTDTFLRKIHAL  348

Qy  600  CAEHLRKYN-----EALQINDTIRMIDAYTHLE-----TFYNEEKDKKFAVIEDDSDE   647
              ||||:.:|     |:| .:|.  ::.|| :||:      | |||:|:|||||||:||
Db  349  CEEHFSPASLDLKFVTPKVIKLLEILRKYKPYERQQFPSVEWTNNRNQDNYVSWSDSED-  407

Qy  648  GGDDEYCDGDEDDLKKPLKLDETDRFLMTLFENNKMLKRLAENPEYENEKLTKLRNT     707
              ||| |    | :|:::    |:|  :      |||::|::|.|| ||  :
Db  408  --DDE----DEEIEEKEKP------ETNFPSPFTNI                        431

Qy  708  IMEQYTRTEESARGIIFTKTRQSAYALSQWITENEK----FAEVGVKAHHLIGAGH--    759
              :|: ||:::|||||||  | .:::|| | |::|:|::     ||:|.||::||||
Db  432  LC----------GIIFVERRYTAVVLNRLIKEAGKQDPELAYISSNFITGHGIGKNQPR  480

Qy  760  ----SSEFKPMTQNEQEKVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIA  815
                  ::|:|  ||.|:|::::||| | ||||||:: |||| ||:|||:|:||:|.|| :|
Db  481  NKQMEAEFR----KQEEVLRKFRAHETNLLIATSIVEEGVDIPKCNLVVRFDLPTEYRS   535

Qy  816  MVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKM-MYKAIHCVQNMK          865
              :||::|||| |.| ..|||::.||.|   ||:|:  :  :||||..|| ::
Db  536  YVQSKGRARAPISNYIMLA------DTDKIKSFEEDLKTYKAIEKILRNK           579
```

RESULT 5

```
ID   DICE_MOUSE               STANDARD;      PRT;   1906 AA.
AC   Q8R418;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Endoribonuclease Dicer (EC 3.1.26.-) (Double-strand-specific
DE   ribonuclease mDCR-1).
GN   DICER1 OR DICER OR MDCR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21886641; PubMed=11889553;
RA   Nicholson R.H., Nicholson A.W.;
RT   "Molecular characterization of a mouse cDNA encoding Dicer, a
RT   ribonuclease III ortholog involved in RNA interference.";
RL   Mamm. Genome 13:67-73(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RL   DNA Res. 9:99-106(2002).
RN   [5]
RP   SEQUENCE OF 1238-1912 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=99162526; PubMed=10051563;
RA   Provost P., Samuelsson B., Radmark O.;
RT   "Interaction of 5-lipoxygenase with cellular proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:1881-1885(1999).
CC   -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
CC       interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
CC       (siRNAs) which target the selective destruction of homologous
CC       RNAs.
CC   -!- SIMILARITY: Belongs to the helicase family.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   -!- SIMILARITY: Contains 1 PAZ domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; AB028449; BAA78691.1; ALT_INIT.
DR   EMBL; AJ132261; CAB38857.2; -.
DR   EMBL; AB023145; BAA76772.2; ALT_INIT.
DR   Genew; HGNC:17098; DICER1.
DR   MIM; 606241; -.
DR   GO; GO:0005622; C:intracellular; NAS.
DR   GO; GO:0003725; F:double-stranded RNA binding; IDA.
DR   GO; GO:0004525; F:ribonuclease III activity; IDA.
DR   GO; GO:0030423; P:RNA interference, targeting of mRNA for des. . .; IEP.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR000999; RNase_III.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00358; DSRM; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS00517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 1.
KW   Helicase; ATP-binding; RNA-binding; Hydrolase; Nuclease; Endonuclease;
KW   Repeat.
FT   NP_BIND     34     41   ATP (POTENTIAL).
FT   SITE       165    168   DECH BOX.
FT   DOMAIN     881   1032   PAZ.
FT   DOMAIN    1266   1393   RNASE III 1.
FT   DOMAIN    1656   1814   RNASE III 2.
FT   DOMAIN    1839   1904   DRBM.
FT   CONFLICT    65     80   VLLTKELSYQIRGDFS -> STTLKKSCLYLDLGETSA
FT                           (IN REF. 1).
FT   CONFLICT   179    179   I -> F (IN REF. 1).
FT   CONFLICT   185    185   N -> I (IN REF. 1).
FT   CONFLICT   204    204   C -> W (IN REF. 1).
FT   CONFLICT   208    208   E -> D (IN REF. 1).
FT   CONFLICT   213    213   I -> V (IN REF. 1).
FT   CONFLICT   383    384   QQ -> HS (IN REF. 1).
FT   CONFLICT   482    483   KQ -> NT (IN REF. 1).
FT   CONFLICT   599    599   D -> H (IN REF. 1).
SQ   SEQUENCE 1912 AA;  217627 MW;  996399DB4B074F21 CRC64;
```

```
RC   STRAIN=Czech II;
RA   Svoboda P., Anger M., Stein P., Schultz R.M.;
RT   "Mouse dicer homolog in oocyte and preimplantation embryos.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22414943; PubMed=12526743;
RA   Doi N., Zenno S., Ueda R., Ohki-Hamazaki H., Ui-Tei K., Saigo K.;
RT   "Short-interfering-RNA-mediated gene silencing in mammalian cells
RT   requires Dicer and eIF2C translation initiation factors.";
RL   Curr. Biol. 13:41-46(2003).
CC   -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
CC       interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
CC       (siRNAs) which target the selective destruction of homologous
CC       RNAs.
CC   -!- TISSUE SPECIFICITY: Expressed in a wide variety of tissues.
CC   -!- SIMILARITY: Belongs to the helicase family.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   -!- SIMILARITY: Contains 1 PAZ domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AF484524; AAL84638.1; -.
DR   EMBL; AF430845; AAM21495.1; ALT_INIT.
DR   EMBL; AB081470; BAC15765.1; -.
DR   MGD; MGI:2177178; Dicer1.
DR   GO; GO:0005622; C:intracellular; ISS.
DR   GO; GO:0003725; F:double-stranded RNA binding; ISS.
DR   GO; GO:0004525; F:ribonuclease III activity; ISS.
DR   GO; GO:0030423; P:RNA interference, targeting of mRNA for des. . .; ISS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR000999; RNase_III.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00358; DSRM; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS00517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   Helicase; ATP-binding; RNA-binding; Hydrolase; Nuclease; Endonuclease;
KW   Repeat.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   SITE        165    168       DECH BOX.
FT   DOMAIN      881   1032       PAZ.
FT   DOMAIN     1266   1393       RNASE III 1.
FT   DOMAIN     1650   1808       RNASE III 2.
FT   DOMAIN     1833   1898       DRBM.
FT   CONFLICT     97     97       A -> C (IN REF. 2).
FT   CONFLICT    157    157       S -> P (IN REF. 3).
FT   CONFLICT    279    279       H -> Y (IN REF. 2).
FT   CONFLICT    600    600       A -> T (IN REF. 3).
FT   CONFLICT    749    749       E -> D (IN REF. 2).
FT   CONFLICT    827    827       T -> I (IN REF. 2).
FT   CONFLICT    878    878       G -> S (IN REF. 1).
FT   CONFLICT    955    955       Y -> C (IN REF. 3).
FT   CONFLICT    983    983       T -> A (IN REF. 3).
FT   CONFLICT   1080   1080       R -> G (IN REF. 3).
FT   CONFLICT   1100   1100       T -> S (IN REF. 3).
FT   CONFLICT   1326   1326       P -> H (IN REF. 1).
FT   CONFLICT   1609   1609       A -> S (IN REF. 3).
FT   CONFLICT   1850   1850       K -> E (IN REF. 3).
SQ   SEQUENCE  1906 AA;  215766 MW;  230BA9BFC19F3091 CRC64;

Query Match             6.7%;  Score 354;  DB 1;  Length 1906;
Best Local Similarity   24.7%;  Pred. No. 8.2e-10;
Matches  158;  Conservative  92;  Mismatches 213;  Indels 176;  Gaps  29;

Qy    309 RPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKD--HLDKKKASEPGKVIVLVNKV 366
              || ||  |    |||  ||| || |||||     |        |   :.:  ||  |
Db     35 RKYQVELLEAALD-HNTIVCLNTGSGKTFIAVLLTKELAHQIRGDLNPHAKRTVFLVNSA 93

Qy    367 -LLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISF----PEVVKSCDIIISTAQIIEN 419
               ||| |  |  || |  ||  ||| | | || |      |  || |:|    ||| |
Db     94 NQVAQQVSAVRTHSDLK------VGEYSDLEVNASWTKERWSQEFTKH-QVLIMTCYV-- 144

Qy    420 SLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVI 479
          : |  ||     |  || |||||||  ||:  | ||       |  |   :
Db    145 ALTVLKNGY---LSLSDINLLVFDECHLAILDHPYREIMK-----LCESCP-- 187

Qy    480 PLPQIGLTASPGVGGATKQAKAEBHILKL--CANLDAFTIKTVKENLDQLKNQIQEPCK 537
          :|:| | |   || ||  | |     | |     | ||  |: || | :  ||   | :
Db    188 SCPRRILGLTASI-LNGKCDPEELEEKIQKLERILRSDAET---ATDLVVLDRYTSQPCE 242

Qy    538 KFAIAD---ATREDPFKEKLLEI---------------MTRIQTYCQMSPMSD-------- 572
          | | |        | |   | |                :.  |  |  || |
Db    243 --IVVDCGPFTDRSGLYERLLMELEAALDFINDCNVAVHSKERDSTLISKQIISDCRAVL 300

Qy    573 --FGTQPYEQWAIQMEKKAAK--KGNRKE--------------------RVCAEHLR---- 605
               |  | ||  |  | | |   ||  |                      || | |
Db    301 VVLGPWCADKVAGMMVRELQKYIKHEQEELRRKFLLPTDTLLRKIHALCEEYFSPASLDL 360

Qy    606 KY--NEALQINDTIRMIDAYTHLE----TFYNEEKDKKFAVIEDDSDEGGDDEYCDGDED 659
                  |||  |:||   || |      | ||  | | |     :|| :  || ::
Db    361 KYVTPKVMKLLEILRKYKPYRRQQFESVEWYNNRNQDNYVSWSDSEDDDDEIEEKKKP 420

Qy    660 EDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESA 719
           :||| | | |    | ||  || |  |    ||| | | |   || || |      |
Db    421 ETNFPSP---------------FTN---------------------------------IL 432

Qy    720 RGIIFTKTRQSAYALSQWITENEK------FAEVGVKAHHLIGAGH------SSEFKPMT 767
          |||| |  | | |  |  | | |        | | || | || ||          | | |
Db    433 CGIIFVERRYTAVVLNRLIKEAGKDDPELAYISSNFITGHGIGKNQPRSKQMEAEFR--- 489

Qy    768 QNEQKEVISKPRTGKINLLIATTVAEEGLDIKECNIVRYGLVTNEIAMVQARGRARADE 827
           |:||:| |:|  |||| |: ||| |   ||   | | | | |  :| ||  ||||| |
Db    490 --KQEEVLRKFRAHETNLLIATSVVEEGVDIPKCNLVVRFDLPTEYRSYVQSKGRARAPI 547

Qy    828 STYVLVAHSGSGVIEHETVNDFREKM-MYKAIHCVQNMK 865
          |  | ||  |      ::  |   |      || |  |
Db    548 SNYVMLA------DTDKIKSFEEDLKTYKAIEKILRNK 579
```

RESULT 6
```
DCR1_CAEEL
ID   DCR1_CAEEL     STANDARD;      PRT;   1845 AA.
AC   P34529;
DT   01-FEB-1994 (Rel. 28, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Endoribonuclease dcr-1 (EC 3.1.26.-).
GN   DCR-1 OR K12H4.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
```

```
RX   MEDLINE=9150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=21521222; PubMed=11641272;
RA   Ketting R.F., Fischer S.E.J., Bernstein E., Sijen T., Hannon G.J.,
RA   Plasterk R.H.A.;
RT   "Dicer functions in RNA interference and in synthesis of small RNA
RT   involved in developmental timing in C. elegans.";
RL   Genes Dev. 15:2654-2659(2001).
CC   -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
CC       interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
CC       (siRNAs) which target the selective destruction of homologous
CC       RNAs. Seems to process the precursor of the small temporal RNA
CC       let-7 which is involved in developmental timing.
CC   -!- SIMILARITY: Belongs to the helicase family.
CC   -!- SIMILARITY: Contains 2 RNase III domains.
CC   -!- SIMILARITY: Contains 1 PAZ domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; L14331; AAA28101.2; -.
DR   WormPep; K12H4.8; CE25057.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR005034; DUF283.
DR   InterPro; IPR0001650; Helicase_C.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR0000999; RNase_III.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF03368; DUF283; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00358; DSRM; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS00517; RNASE_3_1; 1.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   Helicase; ATP-binding; RNA-binding; Hydrolase; Nuclease; Endonuclease;
KW   Repeat.
FT   NP_BIND      33     40       ATP (POTENTIAL).
FT   DOMAIN      759    915       PAZ.
FT   SITE        145    148       DECH BOX.
FT   DOMAIN     1316   1524       RNASE III 1.
FT   DOMAIN     1578   1740       RNASE III 2.
FT   DOMAIN     1768   1831       DRBM.
SQ   SEQUENCE   1845 AA;  210922 MW;  4A96EA2922FF1D9A CRC64;
```

```
Query Match             6.6%;  Score 350;  DB 1;  Length 1845;
Best Local Similarity   23.5%;         Pred. No. 1.2e-09;
Matches  175;  Conservative 118;  Mismatches 249;  Indels 202;  Gaps  37;
```

```
Qy    309 RPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDH------LDKKKASEPGKVIV 361
              :  :   : :     |: |: |  :    :   : |||           : :  ||
Db     14 RDYQVELLDKATK-KNTIVQLGTGSGKTFIAVLLKEYGVQLFAPLDQGGK----RAFF 67

Qy    362 LVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDT-----QLKISFPEVVKSCDIIISTAQI 416
          :|| |||  :    :: :   : : :      |||             :: ||||| :|
Db     68 VVEKVNLVEQ---QAIHIEVHTSFKVGQVHGQTSSGLWDSKEQCDQFMKRHHVVVITAQC 124

Qy    417 LENSLLNLENGEDAGVQLSDFSLIIIDECHHT-NKEAVYNNIMRHYIMQKLKNNRLKKEN 475
          :|   ||       |::::| :::|||||||:     ::   |:         :|||||
Db    125 L-----LDLIRHAYLKIEDMCVLIFDECHHALGSQHPYRSIMVDY--------KLLKKD 170

Qy    476 KPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEP 535
          |||:|        |||     | ::     ||  :      ||| ||    ::  :  :|
Db    171 KPV--PRVLGLTASL-IKAKVAPEKLMEQLKKLLESAMDS-VIETASD-LVSLSKYGAKP 224

Qy    536 ------CKKFAIA--DATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQ 576
                 || |||     || |  ||: |: |   :   |   |  |  |
Db    225 YEVVIICKDFEIGLGIPNFDTVIEIFDETVAFVNTTEFHPDLDPRRPIKDSLKTTR 284

Qy    577 -------PYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQIND-TIRMID-AYTHLE 627
                  :     | :     : | : |:  |:|| : :|: |   : |:   |  :
Db    285 AVFRQLGPWAAW----------RTAQVWEKELGKIIKSQVLPDKTLRFLNMAKTSMI 331

Qy    628 TFYNEEKDKKFAVIEDDSDEGGDEYCDGDEDEDLKKPLKLDETDRFLMTLFFENNKML 687
             |             |   |       :    ||:  |: |  ::    | :
Db    332 TI------KRLL----------------------EPEMKKIKSIEALRPYVPQRVIRLFEIL 365

Qy    688 KRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYAL-----SQWITE 740
          |:|| ||  : | :| || | :| ||  :    ||| : | |  :  ::
Db    366 ETF---NPEFQKERM-KLEKA------------EHLSAIIFVDQRYIAYSLLMMRHIKSW--- 410

Qy    741 NEKFAEVGKVKAHHLIGAG---HSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGL 796
          :|| : |  |  |: || |      :| :  |::|  ||: |  || |||:|::|  ||
Db    411 EPKFK--FVNPDYVVGASGRNLASSDSQGL-HKRQTEVLRRFHRNEINCLIATSVLEEGV 467

Qy    797 DIKECNIVIRYGLVTNEIAMVQARGRARADESTYLVAHSG----------SGVIE 842
          |:|:|||:|| |||| ||  ||  :|| |||||:||  ||           |  :
Db    468 DVKQCNLVIKFDRPLDMRSYVQSKGRARRAGSRYVITVEEKDTAAYCSKLPSDIFTRLVP 527

Qy    843 HETVNDFREKMMYKAIHCVQNMKP--EEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP 900
            |:     :    |: || | ::    ::  |: | : | |: |  :  |  || | :
Db    528 HNQIPIEENGVTK--YCAELLPINSPIKHAI---------------VLKNPMPNKKT----- 569

Qy    901 SLITFLCKNCSVLACSGE-DIHVIEK------------MHHVNMTPE----------FKE 937
          :: || |: | |: | |  |:||:|:              :: : |:            |:
Db    570 AQMAVALEACRQLHLEGELDDNLLPKGRESIAKLLEHIDEEPDEYAPGIAAKVGSSKRKQ 629

Qy    938 LYIVRENKAL-----QKKCADY 954
          :||  |::|      ||| |||
Db    630 LYDKKIARALNESFVEADKECFIY 653
```

```
RESULT 7
YIS2_YEAST
ID   YIS2_YEAST       STANDARD;       PRT;   993 AA.
AC   P40562;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative ATP-dependent RNA helicase YIR002C.
GN   YIR002C OR YIB2C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
```

```
Qy   540  AIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQM-----------EKKAA  590
                    :          ||||  :   :::
Db   285  EV---------PLLLEIEDIIE--------QLGMAVKPVLQQAIELGIYEECDPSQINAF  327

Qy   591  KKGNRKERVCA-----EHLRKYNEAL-----QINDTIRMIDAYTHLETFYNEEKDKKFAV  640
                ||||   ||          :::   ::  ::|: ||| : :::::
Db   328  KAMQQSQKIIANPTIPEGIKWRNFFILQLLNNVGQMLKRLKIY-GIRTFFNYFQNK-----  382

Qy   641  IEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFEN-NKMLKRLAEN-----PE  695
                    :    | ::::::     |||| :     ||||:|  ::: :  ||
Db   383  -----------CTEFTTKYNLKK-----STNKIAAEFYHPILKNIKNQCENYLSDPK  424

Qy   696  YENE-KLTKLRNTIMEQVTRTESARGIIFTKTRQSAYALSQWI--------------------  738
                || ||  :| :: :| |||||  |||||: ::: :
Db   425  FVGHGKLQCVRDELMDFFQKRGSDSRVIIFTELRESALEIVKFIDSVADDQIRPHIPIGQ  484

Qy   739  --------------------TENEKFAEVG------------------------VKAHHLIGAGH  759
                              |||:|: :                      |:|::|:|:
Db   485  ARAKEGFDEVKTRKHAPKGRKKVERLHRQEQKFLEAERTKRAANDKLERSARRTGSSE  544

Qy   760  SSEFKPMTQNEQKEVISKFRTGKINLLIATTVABEGLDIKECNIVIRYGLVTNEIAMVQA  819
                :::|: |: |:|  |:|: : ||||| |||||||||:||||:||||||  :::|:|
Db   545  EAQISGMNQXMQKEVIHNFKKGEYNVLVCTSIGEEGLDIGEVDLIICYDTTSSPIKNIQR  604

Qy   820  RGR-ARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMK-----------PE-  867
                |||   ||||| ::|:  ::::: ::| ||:|::::  ::
Db   605  MGRTGRKRDGKIVLLFSSNESYKFERAMEDYSTLQALISKQCIDYKKSDRIIPEDIIPEC  664

Qy   868  --------------------EYAHKILELQMQSIMEKKMKTKRNTAK  894
                              ||: ::  ::| : ||  :::| |
Db   665  HETLITITINDENEIINEMEDVEVIRYATQCMMGKKVKPKKAITK  708
```

```
RESULT 8
YQHH_BACSU
ID   YQHH_BACSU     STANDARD;      PRT;    557 AA.
AC   P54509;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical helicase yqhH.
GN   YQHH OR BSU24580.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=97124195; PubMed=8969508;
RA   Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA   Kobayashi Y.;
RT   "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT   the Bacillus subtilis genome containing the skin element and many
RT   sporulation genes.";
RL   Microbiology 142:3103-3111(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Duesterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
```

```
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95282515; PubMed=7762303;
RA   Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,
RA   Schwager C., Zimmermann J., Sander C., Ansorge W.;
RT   "Nucleotide sequence and analysis of the centromeric region of yeast
RT   chromosome IX.";
RL   Yeast 11:61-78(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313266; PubMed=9169870;
RA   Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA   Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA   Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA   Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA   Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL   Nature 387:84-87(1997).
CC   -!- SIMILARITY: Belongs to the DEAD box helicase family. DEAH
CC       subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X79743; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z38062; CAA85204.1; -.
DR   PIR; S48436; S48436.
DR   GermOnline; 139714; -.
DR   SGD; S0001441; MPH1.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003724; F:RNA helicase activity; IMP.
DR   GO; GO:0006281; P:DNA repair; IMP.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_box.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   Hypothetical protein; ATP-binding; RNA-binding; Helicase.
FT   NP_BIND   107    114   ATP (POTENTIAL).
FT   SITE      209    212   DEAH BOX.
SQ   SEQUENCE   993 AA;  114057 MW;  474DDC99C543171F CRC64;
```

```
Query Match           5.7%;  Score 303.5; DB 1; Length 993;
Best Local Similarity 20.3%;  Pred. No. 1e-07;
Matches 143; Conservative 118; Mismatches 246; Indels 197; Gaps 25;

Qy   307  QLRPYQMEVAQPALEGKNIIICLPTGSSGKTRVAVVIAKDHLDKKKKASEPGKVIVLVNKV  366
                :|    :: |: :|| :|::|  |: ::  :: |   |||||||  ||||||||||||
Db    86  EVRDYQYTIVHKSL-FQNTLCAIPTGMGKTFIASTVMLNVFRWTKKA------KIIFTAPTR  140

Qy   367  LLVEQLFRKFQPFLKKWYRVIGLSGD-----TQLKISFPEVVKSCDIIISTAQILENS  420
                ||:     ||||||  : ||:|:|       |:|:: :: |: :||:::::||| :
Db   141  PLVAQQ----------IKACLGITGIPSDQTAILLDKSRKNREEIWANKRVFFATPQVVEND  192

Qy   421  LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP  480
                |||  :|  :  ::| || ||||||||: :|: ::: : : |: :::||: :|  :|
Db   193  L------KRGVLDPKDIVCLVIDEAHRATGSSAYTNVKF--------IDRPNSSY--  234

Qy   481  LPQIIGLTASPGVGATKQAKAEEHLLKLCANLDAPTIK-TVKENLDQLKNQIQEPCKKP  539
                |   :|::|| |:|:| |:|  |::|:: :||  | :|  :::||::::|| ::::||
Db   235  --RLLALTATP------ASDLBGVQEVVNNLDISKIERTEESMDIVKYMKKRKKEKI  284
```

RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

DR   EMBL; D84432; BAA12545.1; -.
DR   EMBL; Z99116; CAB14389.1; -.
DR   PIR; G69958; G69958.
DR   Subtilist; BG11699; yqhH.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; Helicase; ATP-binding; Complete proteome.
FT   NP_BIND   83   90   ATP (BY SIMILARITY).
FT   SITE   175   178   DEAH BOX.
SQ   SEQUENCE   557 AA; 64672 MW; 869636CC219081A CRC64.

Query Match   4.2%; Score 223; DB 1; Length 557;
Best Local Similarity   21.5%; Pred. No. 0.00045;
Matches 141; Conservative 98; Mismatches 198; Indels 218; Gaps 31;

Qy   304 PELQLRPYQMEVAQPALEGKN--IICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIV 361
             : :::|||| : |||||| :   : |::   : :: | :::
Db   56 PSFTPLPHQLEVAQKVVEKMNGKAILADEVGLGKTVEAGLIIKEYMIR----GLAKKILI 111

Qy   362 LVNKVLLVEQLFRKEFQPFL-----KKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQ 415
             : | || : | |::   : :::|:
Db   112 LV--PASIVSQWVKELQKFLIPAVEQKKSY------------------VWEQCDIVVSSID 153

Qy   416 ILENSLINLENGEDAGVQLS-DFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKE 474
                        || || |:   :: | :: |||
Db   154 TAKRS------PHREIVLSIPYDLVIIDEAH--------------------KLKNSKTK-- 186

Qy   475 NKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQE 534
             : : || : | | : |: : || | ||| ||
Db   187 --NYBFVRNLVKK 197

Qy   535 PCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGN 594
                 | : |:   : : | ::: | : ||
Db   198 YC---LILTAT--PIQNRIEEIFNLVSL--LKP-GHLGSQNHFQ-----EEFAKKKSS 242

Qy   595 RKERVCAEHLRKYNEALQINDTIRMDA---------YTHLETFYNEEKDKFAVIEDDSDEGG 649
                : | :   : : | || || | : |:   : |
Db   243 LEAH---EHLKDLVNKVMIRN--RRHDTGLNWKQRHVETVPIQFSPSEQALYDEIS---- 293

Qy   650 DDEYCDGDEDEDDLKKPLKLDETDRFLMTL-----------FFENNKML-KRLAENPEYE 697
             :|: || :|| ||| ||| :

Db   294 ----------RLKDSINKPASMFS-----IMTLQRECCSSRBAVYMTLKKMLDQKEKQAPAID 341
Qy   698 NEKLITKLRNTIMEQYRTEESA--------RGIIFTKTRQSAVALSQWITENEKFAE 746
Db   342 EDTISVLIDRI-NQVTQNSKALQVVDLIKKIDDKVIIFTEYRATQIYL-QWFLQQNGISS 399
Qy   747 VGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIR 806
Db   400 V----------PFRGGFKRGKKDWMKDLFRGKIQVLLIATEAGGEGINLQFCNHMIN 445
Qy   807 YGLVINEIAMVQARGRA-RADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHC----- 860
Db   446 YDLPVNPMRLEQRIGRIHRLGQERDV---HIYNNMATKGH-TVEEHILKLLYEKIHLFEKVV 501
Qy   861 ------VQNMKPEEYAHKILELQMQSIMEKKOMKTKRNIAKHYKNNPSLITFL 906
Db   502 GELDDILTKIQVNNFEEHLHDIL---YHSATEEMKIKMD--------NLTSFL 544

RESULT 9
CENE_HUMAN
ID   CENE_HUMAN   STANDARD;   PRT;   2663 AA.
AC   Q02224;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Centromeric protein E (CENP-E protein).
GN   CENPE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93024922; PubMed=1406971;
RA   Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT   "CENP-E is a putative kinetochore motor that accumulates just before
RT   mitosis.";
RL   Nature 359:536-539(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95196755; PubMed=7889940;
RA   Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT   "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT   microtubule motor.";
RL   EMBO J. 14:918-926(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
RN   [4]
RP   FARNESYLATION.
RX   MEDLINE=20459117; PubMed=10852915;
RA   Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA   Bishop W.R., Kirschmeier P.;
RT   "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT   and CENP-F and alter the association of CENP-E with the
RT   microtubules.";
RL   J. Biol. Chem. 275:30451-30457(2000).
CC   -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC   KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC   OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC   AND/OR SPINDLE ELONGATION.
CC   -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC   CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC   QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC   -!- SIMILARITY: Belongs to the kinesin-like protein family.

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z15005; CAA78727.1; -.
DR   PIR; S28261; S28261.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:1856; CENPE.
DR   GK; Q02224; -.
DR   MIM; 117143; -.
DR   GO; GO:0005699; C:kinetochore; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0008350; F:kinetochore motor activity; TAS.
DR   GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR   GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR   GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW   Cell cycle; Centromere; Lipoprotein; Prenylation.
FT   DOMAIN        1    335       KINESIN-MOTOR.
FT   DOMAIN      336   2471       COILED COIL (POTENTIAL).
FT   DOMAIN     2472   2663       GLOBULAR (POTENTIAL).
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
FT   LIPID      2660   2660       S-farnesyl cysteine.
SQ   SEQUENCE   2663 AA;  312087 MW;  CEFC13880C8C8CB8 CRC64;

Query Match             4.1%;  Score 217.5;  DB 1;  Length 2663;
Best Local Similarity   18.6%;  Pred. No. 0.0072;
Matches  209;  Conservative  185;  Mismatches  421;  Indels  307;  Gaps  47;

Qy     9 ENFRYLISCFRARVKMYIQVEPVLDYLITFLPAEVKEQIQRTVATSGNMQAVELLLSTLEK 68
                                                                        :
Db  1040 EQQRKIFSLIQEKNELQQMLESVIAEKEQLKTDLKENIEMTIE--NQEEIRLLGDELKK 1096

Qy    69 GVWHLGWTREFV-----EALRRTGSPLAARYMNPELTDLPSPSFENAHDBYLQLINLLQP 123
                                                                        ::
Db  1097 ------QQBIVAQEKNHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQLLNVQEE 1149

Qy   124 TLVDKLLVRDV--LDKCMEEELLTIE--DRNRIAAAEN-NGNESGVRELLK-RIVQKE-- 175
                :   :::     : :  :
Db  1150 MSEMQKKINEIENLKNELKNELKELTLEHMETERLELAQKLNENYEEVKSITKERKVLKELQ 1209

Qy   176 ------NWFSAFLNVLRQTG------------NNELVQELTGSDCSESNABIENL 212
                                          :: :
Db  1210 KSFETERDHLRGYIRIEIEATGLQTKEELKIAHIHLKEHQETIDELRRS-VSEKTAQIINT 1268

Qy   213 SQVDGPQV------------BEQLLSTTVQPNLEKEVWGMENNSSESSFADSSVVSESDTS 261
              ::
Db  1269 QDLEKSHTKLQEEIPVLHEBQELL------PNVKKVSETQETMNELELLITEQSTTKDS-TT 1322

Qy   262 LAEGSVSCLDESLGHNSNMGSDSGTMGSDSDBENVAARASPEPELQLRPYQMEVAQPALE 321
                :      ::
Db  1323 LARIEM-----ERLRLNEKF------QESQEEIKSLITKERDNLKTIKEALEVKHDQL- 1368

Qy   322 GKNIIICLPTGSGKTRVAVVIAK------------DHLDKKKASEPGKVIVLVNKVLLVEQL 372
                   :  :                    :::: :
Db  1369 ------KEHIRETLAKIQESQSKQEQSLNMKEKDNETTKI-----VSEMEQ- 1408

Qy   373 FRKEFQPFLKKWYRVIGLSGDTQLKISFPE---VVKSCDIIISTAQILENSLLNLENG-- 427
            ::                 :::
Db  1409 FKPKDSALLRIEIEMLGLS--KRLQESHDEMKSVAKEKDDLQRLQEFVLQSESDQLKENIK 1466

Qy   428 EDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIIGL 487
                   :     :::          : :       :  :: :
Db  1467 EIVAKHLETEEELKVAHCCLKEQEETIN------------ELRVNLSEKETE- 1506

Qy   488 TASPGVGATKQAKA------------EEHILKLCANLDAFTIKTVKENLDLKNQIQEPCKKF 539
                                        ::::::
Db  1507 ----ISTIQKQLEAINDKLQNKIQETYEKEEQNIKQISEVQENVNELK-QFKEHRK-- 1558

Qy   540 AIADATREDPFKEKLLEIMTRIQTYCQ-MSPMSDFGTQPYEQWAIQMEKKAAKKGNR- 595
                            :: : :
Db  1559 -AKDSALQSIESKMLELTNRLQESQEBIQIMKEKEEMKRVQEALQIERDQLKENTKEI 1616

Qy   596 ------------KERVC-AEHLRKYNEALQIN-----DTIRMID-AYT 624
                          :: :  :::                  :: ::  :
Db  1617 VAKMKESQEKEYQFLKMTAVNETQEKMCEIBHLKEQFTQKLNLENIETENIRLTQIIHE 1676

Qy   625 HLETFYNEBKDKKPAVIEDDSDEGGDEYCDGDEDEDLKKPLKLD------------ 670
                                                              :
Db  1677 NLEMRSVTKER------DDLRSVEETIKVERDQLKENLRETTRDLEKQEELKIVHMHL 1730

Qy   671 ----ETDRFLMTLFFENNKMLKRLAENPEYENEKL--------TKLRNTIMEQYIRTE 716
                  :   :: ::  :  :::: : :::      : :
Db  1731 KEHQEFIDKLRGIVSEKTNEISNMQKDLEHSNDALKAQDLKIQEELRIAHMLKEQQETI 1790

Qy   717 ESARGIIFTKTRQSAYALSQWITEN---EKPAEVGVKAHHLIGAGHS----------SEFK 764
            :: :    ::  ::  ::   : : :  ::
Db  1791 DKLRGIVSEKTDKLSNMQKDLENSNAKLQEKIQELKANEHQLITLKKDVNETQKKVSEME 1850

Qy   765 PMTQ--NEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGR 822
                    :: :::  :: :  :
Db  1851 QLKKQIKDQSLTLSKLEIENLNI--AQELHENLEEMK---------SVMKERDN 1893

Qy   823 ARADESTYVLVAHSGSGVI-------------EH-ETVNDFREKMMYKAIH 859
              :  :                         ::  :   :      : : :::
Db  1894 LRRVEETLKLERDQLKESLQBTYKARDLEIQQELKTARMLSKEHKETVDKLREKISEKTIQ 1953

Qy   860 CVQNMK------------PEEYAHKILELQMQSIM------------EKKMKTKRNIAKHYKNNPSLI 903
                              ::: :::: :  :::               :::
Db  1954 ISDIQKDLDKSKDELQKDELQLQKKELQLLRVKEDVNMSHKKINEMEQLKKQFEPN---- 2009

Qy   904 TFLCKNCSVLACSGEDIHVIEKMHHVNMTPFKELYIVRENK 945
             :::                 ::
Db  2010 -YLCK------CEMDNFQLTKKLHE---------SLEEIRIVAKER 2039
```

RESULT 10
USO1_YEAST

```
ID   USO1_YEAST     STANDARD;      PRT;   1790 AA.
AC   P25386;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Intracellular protein transport protein USO1.
GN   USO1 OR INT1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180-1A;
RX   MEDLINE=91185402; PubMed=2010462;
RA   Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA   Yamasaki M.;
RT   "A cytoskeleton-related gene, uso1, is required for intracellular
RT   protein transport in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 113:245-260(1991).
RN   [2]
RP   SEQUENCE OF 782-1790 FROM N.A.
RA   Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA   Kendrick K.E.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Bai Y., Symington L.S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
```

```
Qy  365 KVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNL  424
                                 :|||  .|:::| |||  ||
Db 1213 ------------TSEEQSNLKK-----SEIDALNLQIKELKKKNE--TNEASLLE----SI 1250

Qy  425 ENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNN------RLKKENK     476
            :|:  .||  :.|  ||  :: ||
Db 1251 KSVESETVKIKELQ----DECNFKEKEV------SELEDKLKASEDKNSKYLELQKESE 1299

Qy  477 PV------IPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKEN        524
         :                              :|
Db 1300 KIKEELDAKTELKIQLEKITNLSK-----AKEKSESELSRLKKTSSEER--KNAEEQ 1350

Qy  525 LDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQ  584
          ::|||:  |:          |||
Db 1351 LEKLKNEIQIKNQAFE----------KERKL-------                     1389

Qy  585 MEKKAAKKGNRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKFAVIED-   643
            ::  :||:|:   :: |
Db 1390 LEDELIRLQNENEL----------KAKEIDNTRSELEKVSLSNDELLEEKQNTIKSLQDE 1439

Qy  644 ----DSDEGGDDEYC---DGDEDEDDLKKPLKLDETDRPLMTLFFENNKMLKRLAENP    694
             :: :: ||
Db 1440 ILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAKV------EEGLKKLEEES 1493

Qy  695 -----EYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEV    747
            ||       : :|: |: :
Db 1494 SKEKAELEKSKEWMKKLESTIESNETELKSSMETI----RKSDEKLEQ----SKKSAEE 1544

Qy  748 GVKAHHLIGAGHSSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIR- 806
             ::|:::
Db 1545 DIK------NLQHEKSDLISR----------INESEKDIEELKSKLRI             1576

Qy  807 ----YGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIH     859
              :  ||:|: : |
Db 1577 EAKSGSELETVKQELNNAQEKIRINAEENT----------VLKSKLEDIERELKDKQAE 1625

Qy  860 CVQNMKPEEYAH--KILELQMQSIMEKKMKTK  889
             :: ||| ||
Db 1626 IKSNQEEKELLTSRLKELEQELDSTQQKAQKSE 1658
```

RESULT 11
GCC2_HUMAN
ID   GCC2_HUMAN   STANDARD;   PRT;   1583 AA.
AC   Q8TWJ2; O15045; Q8TDH3; Q9H2G8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   GRIP and coiled-coil domain-containing protein 2 (Golgi coiled coil
DE   protein GCC185) (CTCL tumor antigen sel-1) (CLL-associated antigen
DE   KW-11).
GN   GCC2 OR KIAA0336.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
RN   [2]
RP   SEQUENCE OF 1-448 FROM N.A. (ISOFORM 2).
RX   MEDLINE=22188422; PubMed=12200376;
RA   Krackhardt A.M., Witzens M., Harig S., Hodi F.S., Zauls A.J.,
RA   Chessia M., Barrett P., Gribben J.G.;
RT   "Identification of tumor-associated antigens in chronic lymphocytic

CC   -!- FUNCTION: Required for protein transport from the ER to the Golgi
CC   complex.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC   MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC   ER AND THE GOLGI COMPLEX.
CC   -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed
CC   of an heptapeptide repeat pattern characteristic of alpha-helical
CC   coiled coils. May form filamentous structures in the cell.
CC   -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X54378; CAA38253.1; -.
DR   EMBL; L03188; AAB00143.1; -.
DR   EMBL; U53668; AAB66659.1; -.
DR   GermOnline; 140300; -.
DR   SGD; S0002216; USO1.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR006955; Uso1_p115_C.
DR   InterPro; IPR006954; Uso1_p115_head.
DR   Pfam; PF04871; Uso1_p115_C; 1.
DR   Pfam; PF04869; Uso1_p115_head; 1.
KW   Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT   DOMAIN   1   724   GLOBULAR HEAD.
FT   DOMAIN   725   1790   COILED COIL (POTENTIAL).
FT   DOMAIN   465   487   CHARGED (HYPER-HYDROPHILIC).
FT   DOMAIN   991   1790   DISPENSABLE FOR THE PROTEIN FUNCTION.
FT   DOMAIN   1172   1786   ASP/GLU-RICH (ACIDIC).
FT   CONFLICT   847   847   G -> E (IN REF. 2).
FT   CONFLICT   924   924   E -> K (IN REF. 2).
FT   CONFLICT 1253   1253   V -> I (IN REF. 2).
FT   CONFLICT 1319   1319   I -> V (IN REF. 2).
FT   CONFLICT 1461   1461   N -> S (IN REF. 2).
FT   CONFLICT 1581   1581   G -> S (IN REF. 2).
FT   CONFLICT 1600   1600   I -> V (IN REF. 2).
FT   CONFLICT 1661   1661   R -> S (IN REF. 2).
FT   CONFLICT 1772   1772   D -> DEDDDE (IN REF. 2).
SQ   SEQUENCE 1790 AA; 206424 MW; 6CE2B216F9FD4818 CRC64;

Query Match   4.0%;
Best Local Similarity 19.3%;   Pred. No. 0.0058;
Matches 180; Conservative 158; Mismatches 336; Indels 259; Gaps 37; Length 1790;

```
Qy   38 LPAEVKEQIQRTVATSGNMQAVELLLSLTLEKGVWHLGWTREFVEALRRTGS----------  88
            :. :: |::.:|  : ::|
Db  904 LKEDIAAKITEIKAINENLEEMKIQCNNLSKEKEHI--SKELVEYKSRFQSHDNLVAKLT 961

Qy   89 ---PLAARY--MNPELTDLPSPSFENAHDEYLQLLNLLQPTIVDKL------LVRDV    134
                 ::
Db  962 EKLKSLANNYKDMQAENESLIKAVEESKNESSIQLSNL--QNKIDSMSQEKENFQIERGS 1019

Qy  135 LDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAFLNVLRQ-----T   188
              :::
Db 1020 IEKNIEQLKKTISDL----------EQTKEEIISKSDSSKDEYESQISLLKEKLETATT 1068

Qy  189 GNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMENNSSESS  248
              :  ||:.|  :: :| :
Db 1069 ANDENVNKI--SELTKTREELE-----AELAAYKNLKNEL---ETKLETS          1108

Qy  249 FADSSVVSESDTSLARGSVSCLDESLGHNSNMGSDSGTMGS-DSDEENVAARASPEPELQ  307
              :::
Db 1109 EKALKEVKENEHLKEKKIQLKEATETKQQLNSLRANLESLEKEHEDLAA---------Q 1160

Qy  308 LRPYQMEVAQPALEGKNIIICLP--TGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVN   364
              : :: :: : :
Db 1161 LKKYEEQIANKERQYNEEISQLNDEITSTQQENESIKKKKNDELEGEVKAMKS------- 1212
```

RT   leukemia by SEREX.";
RL   Blood 100:2123-2131(2002).
RN   [3]
RP   SEQUENCE OF 1-349 FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abranson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 60-840 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Testis;
RX   MEDLINE=21143360; PubMed=11149944;
RA   Eichmueller S., Usener D., Dummer R., Stein A., Thiel D.,
RA   Schadendorf D.;
RT   "Serological detection of cutaneous T-cell lymphoma-associated
RT   antigens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=22450606; PubMed=12446665;
RA   Luke M.R., Kjer-Nielsen L., Brown D.L., Stow J.L., Gleeson P.A.;
RT   "GRIP domain-mediated targeting of two new coiled-coil proteins,
RT   GCC88 and GCC185, to subcompartments of the trans-Golgi network.";
RL   J. Biol. Chem. 278:4216-4226(2003).
CC   -!- FUNCTION: Probably involved in maintaining Golgi structure (By
CC   similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; peripheral membrane protein
CC   associated with the trans-Golgi network.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=Q8IWJ2-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=Q8IWJ2-2; Sequence=VSP_007733;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC   -!- SIMILARITY: Belongs to the golgin family.
CC   -!- SIMILARITY: Contains 1 GRIP domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AB002334; BAA20794.1; -.
DR   EMBL; AAL99918.1; AAL99918.1; ALT_INIT.
DR   EMBL; BC037774; AAH37774.1; -.
DR   EMBL; AF273042; AAG34902.1; ALT_INIT.
DR   Genew; HGNC:23218; GCC2.
DR   InterPro; IPR000245; GRIP_domain.
DR   InterPro; IPR003345; M_repeat.
DR   InterPro; IPR002017; Spectrin.

DR   Pfam; PF01465; GRIP; 1.
DR   Pfam; PF02370; M; 9.
DR   PROSITE; PS50913; GRIP; 1.
KW   Golgi stack; Coiled coil; Alternative splicing; Polymorphism.
FT   DOMAIN   9   1517   COILED COIL (POTENTIAL).
FT   DOMAIN   138   947   GLU-RICH.
FT   DOMAIN   1508   1558   GRIP.
FT   VARSPLIC   1   1   M -> MEDLVQDGVASPATPGTGKSKLETLPKEDLIKFAKK
FT   QMMLIQKAKSRCTELEKEIEELRSKPVTEGTDIIKALTER
FT   LDALLEKAETEQCLSLKKENIKM (in isoform 2).
FT   /FTId=VSP_007733.
FT   VARIANT   1197   1197   R -> G (in dbSNP:1061202).
FT   /FTId=VAR_016101.
FT   CONFLICT   593   599   NNKLSSE -> TINSVQK (IN REF. 2).
FT   CONFLICT   836   836   S -> T (IN REF. 2).
SQ   SEQUENCE   1583 AA; 184657 MW; 0FAB4B7E5969111 CRC64;

Query Match   3.9%; Score 205.5; DB 1; Length 1583;
Best Local Similarity   17.5%; Pred. No. 0.014;
Matches 195; Conservative 215; Mismatches 407; Indels 297; Gaps 52;

Qy   42 VKEQIQRTVATSGNMQAVELLLSTLEKGVWHLGWTREFVEALRRTGSPIAARY------   94
           :|::|:|.|| |:|:   ::| :|::|:| |:.:|:|:
Db   1 MKQEVEDSVTKMGDAH----------------KELEQS--HINYVKE-IENLKNELMAVRSKYSEDKANL   51

Qy   95 ------------MNPELTDLPSPSFENAHDEYLQLL---NLLQPTLVDKLL-VRDVLDKCMEEE 142
                       :| .| |:|:| ::|:
Db   52 QKQLEEAMNTQLELSEQLKFQNNSEDNVKKLQEIEKIRPGFEQILYLQKQLDATTDEK 111

Qy   143 LLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAFLNVLRQTGNNELVQELTGSDC 202
           ||.:|:|:|| :.| | | |:|:|::| | | |::|:|:|
Db   112 KETVTQLQNIEANSQHYQKNINSLQEELLQ---------LKAIHQEEVKELMCQIEAS-A 162

Qy   203 SESNAEIENLSQVDGPQVEE-QLLSTTVQPNLEKEVWGMENNSS-------------ES 247
           ||:|:||:| |:| | |:|:|:| |::|
Db   163 KEHEAEINKLNELKENLVKQCEASEKNIQKKYECEILENLRKATSNANQDNQICSILLQEN 222

Qy   248 SFADSSVVSESDTSL---AEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARAS 301
           | :.:|:|:.|| .|: :.|| ::|. | | | || |. |::|
Db   223 TFVE-QVVNEKVKHLEDTLKELESQHSILKDEVTYMNNL----KLKLEMDAQHIKDEFF 276

Qy   302 PEPE-LQLRPYQM---------EVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDK 349
           | |. ||.| .:.:|:.||| | ||| |.|| | | |::|||:|
Db   277 HEREDLEFKINELLLAKEEQGCVIEKLKSELAGLNKQFC-------------YTVEQHNRE 324

Qy   350 KKKASEP-GKVIVLVNKVLLVEQLFRKEFQPPLKKWYRVIGLSGDTQ-LKISFPEVVKSC 407
           ||:|:| ||||:|:|||:|| .|:..|.|:: ||:||||| ||:||| |:|:|
Db   325 VQSLKEQHQKEISELNETFLSDSEKEK-----LTLMFEIQGLKEQCENLQQEKQEAILNY 379

Qy   408 DIIISTAQILENSLLNLENGDAGVQLSDFSLIIIDECHITNKEAVYNNIMRHYLMQKLK 467
           :: | |:||:|||:|:||| ||:| |:|::|:| |:|:||:|:| | : :||||
Db   380 E---SLREIME--ILQTELGSSAGKISQEF-----------ESMKQQQASDV--HELQQKLR 423

Qy   468 N---------NRLKENKPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDA 515
           | |:|:|:|| |:||| |||| | |.| || :: ||:.| |:||
Db   424 TAFTEKDALLETVNRLQGENEKLLSQQELV----PELENTIKNLQEKNGVYLLSLSQRD 478

Qy   516 FTIKTVKENLDQLKNQIQEPCKKFAIADATRE-DPPKEK------LLEIMTRIQTYCQM 567
           ||:|::|::|.|||:| |:| ||::| ||: |:| ||:| |
Db   479 TMLKELEGKINSLTEKDDFINK-LKNSHEEMDNFHKKCEREERLIELGKKVEQTIQY 536

Qy   568 S---------------PMSDFGTQPYEQWAIQM------------------------- 585
           | | |. |. || :|
Db   537 NSELEQKVNELTGGLEETLKEKDQNDQKLEKLMVQMKVLSEDKEVLSAEVKSLYEENNKL 596

Qy   586 --EKKAAKK-------GNRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEKDKK 637
             :||||| ||:|:| |:|:|: | |:|:| ||::||
Db   597 SSEEKKQLSRDLEVFLSQKEDVILKEHITQLEKKLQL----MVEEQDNLNLLENEQVQK 651

Qy   638 FAV------------IEDDSDEGG-----DDEYCD---GDEDEDDLK 664
           | | :||.|:| | ||:||| |||:||||
Db   652 LFVKTQLYGFLKEMGSEVSEDSEEKDVVNLQAVGESLAKINEEKCNLAFQRDEKVLELE 711

DR   EMBL; U18922; AAB64699.1; -.
DR   PIR; S50675; S50675.
DR   GermOnline; 139249; -.
DR   SGD; S0000974; BRR2.
DR   GO; GO:0005687; C:snRNP U4; IDA.
DR   GO; GO:0005682; C:snRNP U5; IDA.
DR   GO; GO:0005688; C:snRNP U6; IDA.
DR   InterPro; IPR003593; AAA ATPase.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004179; Sec63.
DR   Pfam; PF00270; DEAD; 2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF02889; Sec63; 2.
DR   SMART; SM00382; AAA; 2.
DR   SMART; SM00490; DEXDc; 2.
DR   SMART; SM00487; HELICc; 1.
DR   SMART; SM00611; SEC63; 2.
KW   Hydrolase; Helicase; mRNA processing; mRNA splicing; Spliceosome;
KW   Nuclear protein; ATP-binding; Repeat.
FT   DOMAIN   521   923   I.
FT   DOMAIN   1370   1757   II.
FT   NP_BIND   521   528   ATP (POTENTIAL).
FT   NP_BIND   1370   1377   ATP (POTENTIAL).
FT   SITE   634   637   DEIH BOX.
FT   SITE   1474   1477   DDAH BOX.
SQ   SEQUENCE   2163 AA; 246183 MW; DFAF7E3B7168D944 CRC64;

Query Match   3.9%; Score 204.5; DB 1; Length 2163;
Best Local Similarity   19.1%; Pred. No. 0.024;
Matches 162; Conservative 136; Mismatches 247; Indels 305; Gaps 40;

Qy 127 DKLLVRDVLDKCMEH--------------------------ELLTIED----------RNR------IAAAE 156
         : :::||:|:| | ||| ||:
Db 305 DTSVIQDLSEKILNDIETLEHNPVALEQKLVDLLKFENISLAEFIILKNRSTIFWGIRLAK 364

Qy 157 NNGNESGVRELLKRILVQKENWFSAFLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVD 216
         :|: :| |::::|
Db 365 STENE--IPNLIEKMVAK--------------GLNDLVEQYKFRETTHSKRELD--SGDD 406

Qy 217 GPQVEE----------QLLSTTVQPNLEKEVWGMENNSESSFADSS-VVSESDTSLAEGSVSCLD 271
         || ||| ||| : : ::| : ::::
Db 407 QPQSSEAKTKFSNPAIPPVIDLE----------KIKFDESSKLMTVTKVSLPEGSFK--- 454

Qy 272 ESLGHSNSMGSDSGTMGSDSDEENVAARASPEPELQLR----------------------- 309
Db 455 --------------RVKPQVDEIHIPAPSKPVIDYELKEITSLPDWCQEAFPSSETTSLN 500

Qy 310 PYQMEVAQPALEG-KNIIICLPTGSGKTRVAVY---------IAKDHLDKKKKASEPGKVIVLV 363
         ||| | |:| ||:||:|:|||||||| | | ||| :: ::::| ::
Db 501 PIQSKVFHAAFEGDSNMLICAPTGSGKTNIALLITVLKALSHHYNPKTKKLNLSAFKIVYI 560

Qy 364 N-KVLLVEQLFRKEFQ--PFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILE 418
         : ::::|:|:| :| : |||||:|::|||:| :| ||:
Db 561 APLKALVQEQV--REFQRRLAFL--GIKVAELTGDSRLS--RKQIDETQVLVSTPEKWD 613

Qy 419 NSLLNLENGEDAGVQLSDFSLIIIDECH--HTNKEAVYNNIM-RHYLMQKLKNNRLKKEN 475
         || |: || || ||| ::|:||||| | : :: :::| | : ::: :::
Db 614 ITTRNSNN--LAIVEL--VRLLIDEIHLLHDDRGPVLESIVARTFWASKYGQE------ 663

Qy 476 KPVIPLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEP 535
         : :::| | :|:| | | | : ::|:::| | |:::|::||::|: ::::
Db 664 ----YPRIIGLSAT-------------------------LRVPKEGLFYFDSSFR-P 700

Qy 536 CKKFAIADATREDPFKEKLLEIMTRIQYCQMSPMSDFGTQPYEQWAICMEKKAAKKGNR 595

Qy 665 KPLKLDETDRF----LMTLF--FENNKMLKRLAENPEYENEK------LTKLRNTIMEQ 711
         |:| | :||||:| |:|||: :: |:   :::::||
Db 712 KEIKCLQEESVVQCEELKSLLRDYEQEKVLLR-KELEBIQSEKEA.QSDLLEMKNA--NE 768

Qy 712 YTRTERSARGIIFTKRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSHSSEFKPWTQNEQ 771
         || ||
Db 769 KTRLENQN--LLIQVEEVSQTCSKSEIHNEK-EKCFIKEH------ENLKPLLEQKE 816

Qy 772 ------KEVISKFRTGKINLLIATTVAEEGLD--IKECNIVIRYGLVTNEIAMVQ 818
         :|::|::| || | :|: || : :
Db 817 LRDRRABLILLKDSLAKSPSVKNDPLSSVKELEEKIENLEKECK--EKEBKINKIKLVA 873

Qy 819 ARGRARADESTYVLVAHSGSGVIEHETVNDFREKMYKAIHCVQNMK-----PERYAHKI 873
         | ::   :::||
Db 874 VKAKKELDSSRK--------ETQTVKEELESLRSEKDQLSASMRDLIQGAESYKNLL 922

Qy 874 LELQMQSIMEKKMKTKRNIAKHYKNPLSLIIFLCKN----CSVLACSGRDIHVIEKMHHV 929
         || : :| :|::| ::::| ::|:
Db 923 LEYEKQS--EQLDVEKERANNFEHRIEDLTRQLRNSTLQCETINSDNEDL--LARIETL 977

Qy 930 NMTPEFKELYIVRENKALQKKCADYQINGEIICK 963
         :: ::| |:: |: |
Db 978 QSNAKLLEVQILEVQRA--KAMVDKELEAEKLQK 1009

RESULT 12
BRR2_YEAST
ID   BRR2_YEAST   STANDARD;   PRT;   2163 AA.
AC   P32639;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pre-mRNA splicing helicase BRR2 (EC 3.6.1.--) (Protein Snu246).
GN   BRR2 OR RSS1 OR SNU246 OR YER172C OR SYGP-ORF66.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313264; PubMed=9169868;
RA   Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA   Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA   Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA   Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA   Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA   Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA   Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL   Nature 387:78-81(1997).
RN   [2]
RP   CHARACTERIZATION.
RC   STRAIN=DBY473;
RX   MEDLINE=96304576; PubMed=8722763;
RA   Noble S.M., Guthrie C.;
RT   "Identification of novel genes required for yeast pre-mRNA splicing
RT   by means of cold-sensitive mutations.";
RL   Genetics 143:67-80(1996).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=96324408; PubMed=8670905;
RA   Lauber J., Fabrizio P., Teigelkamp S., Lane W.S., Hartmann E.,
RA   Luehrmann R.;
RT   "The HeLa 200 kDa U5 snRNP-specific protein and its homologue in
RT   Saccharomyces cerevisiae are members of the DEXH-box protein family of
RT   putative RNA helicases.";
RL   EMBO J. 15:4001-4015(1996).
CC   -!- FUNCTION: REQUIRED FOR PRE-MRA SPLICING.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DOMAIN: Composed of two similar domains.
CC   -!- SIMILARITY: Belongs to the helicase family. SKI2 subfamily.

```
FT   NP_BIND      35     42       ATP (POTENTIAL).
FT   DOMAIN      197    415       COILED COIL (POTENTIAL).
FT   DOMAIN      446   1003       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1163 AA;  135507 MW;  CE5F0BD2215D77A92 CRC64;

Query Match     3.7%;  Score 197.5;  DB 1;  Length 1163;
Best Local Similarity  18.8%;  Pred. No. 0.022;
Matches  197;  Conservative 187;  Mismatches 399;  Indels 263;  Gaps  45;

Qy    31 VLDYLTF-LPAEVKEQIQRTVATSGMQAVELLLSLEKGVWHLGWTREFVEALRTGSP   89
         :||:|| |  |: :  ||||
Db    44 ILDSITLSHYGEVARKSSNFMNTNCNSLNVSFEFQISGKEIKRYLVEREF----RRDNKT   99

Qy    90 LAARYMNPELTDLPSPSFENAHDEYLQLLNLLQPTIVDKLLVRDVLDKCMEEELLTIEDR  149
Db   100 GSVRSKSAKIVDITGDEVEVLEEG------------AKSVNEKCQEIIGLSLDDF  142

Qy   150 NRIAAAEN------------NGNESGVRELLKRIVQKENW-----FSAFLNVLRQ-TGNNELV  194
Db   143 TRTVVLPQGKFSEFLKLEGKER--RNMLERLFNLQEYGDELSFKLARKIRKEREKENVLV  200

Qy   195 QELTG------SDCSESNAELENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMENNSSESSF  249
Db   201 GELKGYENINEDVLKERRELLKENNDFFNEASKEYLKABEEYNEGKEYWGLQIEEKNR  260

Qy   250 ADSSVVSBSDTSLAEGSVSCLDESLGHRSNMGGSDSGTMGSDSDEENVAARASPEPELQLR  309
Db   261 VRKDLMEKKD--------------EIDLKEKRARLG-ESSSKVK  289

Qy   310 PY----------QMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKGKKASEPGKVI  360
Db   290 PYIDNYENTLKQIDILKEQILSRB------NTMKAISLEKEDMEKKLSIAKDNKEK  339

Qy   361 VL---VNKVLLVEQLFRKEFQPFLK-KWYRVIGLSGDTQLKISFPEV----VKSCDII  410
Db   340 ALPKFMIKHHIILDAIKEKDLLDNIKLEKKRIQGKIEKLSLEASNKEELIKQNIKDISL  399

Qy   411 ISTAQILENSLLNLENGEDAGVQLSD--FSLIIIDE-CHHTN-----------  449
Db   400 TLKIQNLESKIDNLKVPEEYKNKINEGIFLLRNYDEKLKHKNKLGLDCDKFQVDFEKAKS  459

Qy   450 -KEAVYNNIMRHYLMQKLKNNRLKKKENKPVIPLPQIIGLTASPGVGGA-TKQAKAEEHIL  507
Db   460 KKEMLFNKLEEERSKLDTYTKKLQDLNKD-FPKDDVL-LTFQEKLNDSRQKWAKYSEYNE  517

Qy   508 KLCANLDAF-------TIKTVKENLDQLKNQIQ--------EPCKKPAIADATRE-----  547
Db   518 SLKASLRVVENSEQVLRTKK&EMTKLEDKISKVNIKIESLETENMAHVLREKLKSGEACP  577

Qy   548 --------DPFKEKLLEIMTRIQTYCQMSPMSDF-GTQPYE----------QWAIQMEKKAA  590
Db   578 VCGSVHHIKEGFKEVDLKALETLK-----------SELEGFEKKRKFENEEIVMCEASIKVEEKNI  632

Qy   591 KKGN----------------RKERVCAEHLRKYN-EALQINDTIR-MIDAYTHLETF  629
Db   633 KKLNESINNLGEEFKEVSLESMEKKFNYLKEKVNKFNLEKIQLDDNIKDLSERSNKIEVE  692

Qy   630 YNEEK--------DKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFEN  683
Db   693 YQKEKTVEKQCEKRIVDLKSELEE-AIKEFNEVAYTIENLKAELKIQD------FKFEM  744

Qy   684 NKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEK  743
Db   745 KEILEKERVRVEAEGB-IKDLRNLLNIRHTEKEQ--------LMDKCSRLKEELSKNK-  793

Qy   744 FAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNI  803
Db   794 -AEL--------KEKDKIINEKIELI-KNKVGVLDNLYELKEKIEG-TIKK--I  834

Qy   804 VIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQN  863
Db   835 EEQYNLCDKKMNEIEDKYRKCSDE----IIKYHSNLSSLKDRKVND---------IDKLNK  882

Db   701 C-----------------------------------------------PLS----QQFCGIKERNSLKKLKA  721
Qy   596 KERVCAEH-LRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYC  654
Db   722 MNDACYEKVLESINEGNQI----IVFVHSRKET--------------------------  750
Qy   655 DGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTR  714
Db   751 --------------SRTATWLKNKFAEEN--------------------ITHKLTK  772
Qy   715 TEESARGIIFTKTRQSAYALSQWITENEKRPAEVGVKAHHLIGAGHSSEFKPMTQNEQKEV  774
Db   773 NDAGSKQILKT---EAANVLDPSL--RKLIESGIGTHH---AG-------LTRSDRSLS  816
Qy   775 ISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLV------------TNEIAMVQARG  821
Db   817 EDLFADGLLQVLVCTATLAWGVNLPAHTVIIKGTDVYSPEKGSWEQLSPQDVLQMLGRAG  876
Qy   822 RARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSI  881
Db   877 RPRYD--TF-----GEGII----ITDQSNVQYYLSVLNQQLPIESQFVSKLVDNLNAEV  924
Qy   882 MEKKMKTKRN  891
Db   925 VAGNIKCRND  934
```

```
RESULT 13
SBCC_CLOAB
ID   SBCC_CLOAB     STANDARD;      PRT;   1163 AA.
AC   Q97FK1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nuclease sbcCD subunit C.
GN   SBCC OR CAC2736.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
CC   -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC       can inhibit DNA replication and are intermediates in certain DNA
CC       recombination reactions. The complex acts as a 3'->5' double
CC       strand exonuclease that can open hairpins. It also has a 5'
CC       single-strand endonuclease activity (By similarity).
CC   -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC   -!- SIMILARITY: Belongs to the SMC family. SbcC subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE007771; AAK80682.1; -.
DR   PIR; G97236; G97236.
DR   InterPro; IPR003439; ABC_transporter.
KW   Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW   DNA recombination; ATP-binding; Coiled coil; Complete proteome.
```

Qy   864 MKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVI 923
                : :: :: : :::: .::   ::   : :| ||: :::
Db   883 ILMEEKFENIEKAKENYLNDKEINLLKSDVEKYKNELSKV-----NGAVEVLS------- 930

Qy   924 EKMHHVNMTPEFFKELYIVRENKALQK 949
                :|: :: :: :| ::| :|
Db   931 KKILKNRKLT--EEKWIBIQNNRVEK 953

RESULT 14
GCC2_MOUSE
ID   GCC2_MOUSE   STANDARD;   PRT;   1679 AA.
AC   Q8CHG3; Q8BR44; Q8RQ05; Q9CT45;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   GRIP and coiled-coil domain-containing Protein 2 (Golgi coiled coil
DE   protein GCC185).
GN   GCC2 OR KIAA0336.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 1-769 FROM N.A.
RC   TISSUE=Breast cancer;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE OF 1-442 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Brain, and Embryo;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlested C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itch M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [3]
RP   SEQUENCE OF 49-1679 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
CC   -!- FUNCTION: Probably involved in maintaining Golgi structure (By
CC   similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; peripheral membrane protein
CC   associated with the trans-Golgi network (By similarity).
CC   -!- DOMAIN: Extended rod-like protein with coiled-coil domain.
CC   -!- SIMILARITY: Belongs to the golgin family.
CC   -!- SIMILARITY: Contains 1 GRIP domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; BC027339; AAH27339.1; ALT_INIT.
DR   EMBL; AK011206; BAB27466.2; -.
DR   EMBL; AK045701; BAC32463.1; -.
DR   EMBL; AB093232; BAC41416.1; -.
DR   PIR; PT0649; PT0712.
DR   MGD; MGI:1917547; 2600014C01Rik.
DR   InterPro; IPR000237; GRIP_domain.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF01465; GRIP; 1.
DR   PROSITE; PS50913; GRIP; 1.
KW   Golgi stack; Coiled coil.
FT   DOMAIN   31   1613   COILED COIL (POTENTIAL).
FT   DOMAIN   83   501   GLU-RICH.
FT   DOMAIN   1604   1654   GRIP.
FT   CONFLICT   440   440   I -> M (IN REF. 2).
SQ   SEQUENCE   1679 AA; 194443 MW; 6E2F84263E5E0103 CRC64;

Query Match   3.7%; Score 197; DB 1; Length 1679;
Best Local Similarity 19.2%; Pred. No. 0.039;
Matches 227; Conservative 179; Mismatches 438; Indels 338; Gaps 53;

Qy   19 RARVKMYIQVEPVLDYLTFLPAEVKE---------QIQRTVATSGNMQAVE------LIL 63
               ||:|:||| ::::| ||:||||||| :::|
Db   96 KENVKMKQEVE--DSVTKLEETHKEFEQSHRNVVKEIRSCKNELMAVHSEHSKETAILQ 152

Qy   64 STLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPS--------PSFENAHDE 113
               |||| : :|| ||: :: ||::| :::::| :|
Db   153 KELEEAVHKQVELREQLKS--QSDSEDNVRKLQEEIQNITAAFEEQISCLEKKLEATSDE 210

Qy   114 YLQLLNLLQPTLVDKL--LVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRI 171
               | ::::| ::| || :::|: |:|| ||||||||||| ||:||||| |: ||||||||
Db   211 KQQBIIHLQKVIEDKAQHYQKDI--NTFQAEILQLRATHK---------EEVTELMSQI 258

Qy   172 VQKENWFSAFLNVLRQTGNNELVQELTGSDCSES-----NAEIENLSQV-DGPQVEEQLL 225
               :|:|:   :::| :: : || :::| ::||| ||:| ::| :::|:
Db   259 ETSAKEHEAEINKLKE------NRVTQCEASENIPEKYQCESENLNEVASDASPESQNC 311

Qy   226 STTVQPNLEKEVGMEENNSESSFADSVVSRSDTSLAEGSVSCLDESLGHNSNMGSDSG 285
               :: :::::| : ::: :: :: :| :::| :|

```
Db   312 SVALQ----------EDPSAEQTVCDKVRQLEDSLKELESQHSILKDEVTYMNNL----- 356
                        :||:  : .::|:
Db   286 TMGSDSDEENVAARASPEPE-----------LQLRPYQMEVAQPALEGKNIIICLPTGS 333
         .:  :|: ::::
Db   357 KLKLEMDAQHIKDEFFHEREDLEFKINELLLAKEEQGYVVEKLKYEREDLNRQLCCAVEQ 416
                        |  | .:.
Qy   334 GKTRVAVYIAKDHLDKKKASEPGKVIVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGD 393
                        ||  ||    ::
Db   417 HNKEIQ-RLQEHH--QKEVSE--------LSETFISGSFKEKLALMFEIQGLKEQ 460
         :.:.|  :|||||   ::|
Qy   394 TQ-LKISFPEVVKSCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEA 452
         :: .|  :||  || |: || ::|||  ||: :::||| :|: :|  :| :|::.
Db   461 CENLQHEKQEVVLNYESLREMMEILQTEL-------GESAGKISQEF-----------ETMKQQ 506
                        || |:  |::| :
Qy   453 VVNNIMRHYLMQKLKN--------------NRLKKENKPVIPLPQILGLTASPVGVGGATKQA 500
                        |||:| | : :.:  :.: |: ||
Db   507 QASDV--HELQQKLRSAFNEKPALLETVNRLQGENEKLLSQQELV------PELESTIKNL 559
         : :
Qy   501 KA-----------EEHILKLCANLDAF--TIKTVKENLDQL---------KNQIQ 533
         :|
Db   560 QADNSMYLASLGQKDTMLQELEAKISSLAKEKDDFISKIKTSHEEMDDLHQKWEREQRLS 619
                        :::  :: |:::
Qy   534 EPCKKFAIADATREDPFKEKLLEIMRIQTYCQMSPMSD--------FGTQPYFEQWAIQMEK 587
         :: :|:  ||||  :::  |:|:||
Db   620 VELREAAGQAAQHNSELRQRVSELTGKLDELVREKSQNDQSITVQMKTMTEDQEALSSKI 679
                        ||:.:|: :
Qy   588 KAAKKGNRKERVCAEHLRKYNEALQINDTIRMIDA---YTHLETFYNEEKDKKFAVIEDD 644
                        ||:|:|| ::.: ||  |||||  |: :|:|
Db   680 KSLYEENN---------RLHSEKAQLSRDLEALQAQQDFAHKEHVAEFEKKLQLMV----- 726
         :| :
Qy   645 SDEGGGDDEYCDGDEDEDLKKPLKLDETDR-FLMTLFFENNKMLKR--LAENPBYENEKL 701
                        :: |:  :: |||:| |:
Db   727 ----------EERDDLNKLLENEQVQKSFVKTQLYEYLKQLRASILEEN--EEEDV 770
                        |||  |:| :  :
Qy   702 TKLRNTIMEQYTRTESARGIIFTKTRQSAYALSQWITENEKFAEVGKAHHLIGAGHSS 761
         :: : :|||  :: |: :|||| ||
Db   771 VKLIQAVGESLVKVKEEEHNIVF-----EYDARVLELENK------IKCLQEDSAVQCE 818
                        :: : |: |:  ||   |:
Qy   762 EFKPMTQNERQKEVISKFRTGKINL---LIATTVAEEG--LDIKECNIVIRYGLVTNEIAM 816
                        ||:: ||:| :.| ::::
Db   819 ELRTLVRDSEQE--------KILRKELDAVTSAKEALQLDLLEMK------NTNEKAS 863
         :: : :||||| :|:
Qy   817 VQARG-RARADESTVLVAHSGSG---------VIEHET---------VNDFR-------- 850
         : : |  |||:|
Db   864 LENQTLSTQVEELSQTL--HSRNEVHDEKVLVIEHENLRLILKQRESEIQDVRAELILLK 921
                        ||:
Qy   851 ----------ERMMYKAIHCVQNMKPEYAHKI 873
                        |::
Db   922 DSLEKSPSVKDQLSLVKELEEKIEBSLEKBSKDKDEKISKIKLVAVKAKKELDSNRKEAQT 981
                        ||  ||
Qy   874 LEIQMQSIMEKK----------MKTKRNIAKHYKNMPSLITFLCKNCSVLACSGEDIHVIEKMH 927
                        ||| :||:
Db   982 LREELESVRSEKDRLSASMKEFLQGAESYK---SLLLEYDKQSEQLDVEKERAHNFER-- 1036
                        :|
Qy   928 HV----------NMTPEFKELYIVRENKALQKKCADYQINGEI----ICKCGQAWGTMWVHK 975
                        :.|
Db  1037 HIEDLTKQLRNSTCQYERL--TSDNEDLLARIETLQANAKLLEAQILBVQKAKG--VVEK 1092
         :||
Qy   976 GLDLPCL---KIRNFV--------VVFKNNSTKKQYKKWVE 1005
                        ||:
Db  1093 ELDAEELQKEQKIKEHVSTVNELEELQLQFQKEKKQLQKTMQ 1134
```

RESULT 15
GOA4_HUMAN
ID   GOA4_HUMAN   STANDARD;   PRT;   2230 AA.
AC   Q13439; Q13270; Q13654; Q14436;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Golgi autoantigen, golgin subfamily A member 4 (Trans-Golgi p230) (256
DE   kDa golgin) (Golgin-245) (72.1 protein).
GN   GOLGA4.

OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=96215236; PubMed=8626529;
RA   Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT   "Molecular characterization of trans-Golgi p230: a human peripheral
RT   membrane protein encoded by a gene on chromosome 6p12-22 contains
RT   extensive coiled-coil alpha-helical domains and a granin motif.";
RL   J. Biol. Chem. 271:8328-8337(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Seelig H.P.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 131-2230 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=96125112; PubMed=8537393;
RA   Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT   "Molecular characterization of golgin-245, a novel Golgi complex
RT   protein containing a granin signature.";
RL   J. Biol. Chem. 270:31262-31268(1995).
RN   [4]
RP   SEQUENCE OF 524-672 FROM N.A.
RC   TISSUE=Gastric fundus;
RA   Balague C.;
RL   Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC   -!- FUNCTION: May play a role in vesicular transport from the trans-
CC   Golgi.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
CC   associated with the trans-Golgi network.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=4;
CC   Comment=Additional isoforms seem to exist;
CC   Name=1;
CC   IsoId=Q13439-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
CC   Name=3;
CC   IsoId=Q13439-3; Sequence=VSP_004274;
CC   Name=4;
CC   IsoId=Q13439-4; Sequence=VSP_004275;
CC   -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC   -!- DISEASE: Antibodies against GOLGA4 are present in sera from
CC   patients with Sjogren's syndrome (SS). [MIM:270150]. Sera from
CC   patients with SS often contain antibodies that react with normal
CC   components of the Golgi complex.
CC   -!- DISEASE: Antibodies against GOLGA4 are found in sera from
CC   hepatitis B patients.
CC   -!- SIMILARITY: Belongs to the golgin family.
CC   -!- SIMILARITY: Contains 1 GRIP domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U41740; AAC50434.1; -.
DR   EMBL; X82834; CAA58041.1; -.
DR   EMBL; U31906; AAC51791.1; -.
DR   EMBL; X76942; CAA54261.1; -.
DR   Genew; HGNC:4427; GOLGA4.
DR   MIM; 602509; -.
DR   MIM; 270150; -.
DR   GO; GO:0005802; C:Golgi trans face; TAS.
DR   GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR   InterPro; IPR000237; GRIP_domain.
DR   Pfam; PF01465; GRIP; 1.

```
DR   PROSITE; PS50913; GRIP; 1.
KW   Antigen; Golgi stack; Coiled coil; Alternative splicing.
FT   DOMAIN       133    2185       COILED COIL (POTENTIAL).
FT   DOMAIN       252    2096       GLU-RICH.
FT   DOMAIN      2168    2215       GRIP.
FT   VARSPLIC    2154    2185       TPYKGGNLYHTDVSLFGEPTEFEYLRKVLFEY -> HLTKV
FT                                  AICTIRMSHSLENLPNLSICEKCPLSI (in isoform
FT                                  2).
FT                                  /FTId=VSP_004272.
FT   VARSPLIC    2186    2230       Missing (in isoform 2).
FT                                  /FTId=VSP_004273.
FT   VARSPLIC    2103    2109       Missing (in isoform 3).
FT                                  /FTId=VSP_004274.
FT   VARSPLIC    2222    2230       FTSPRSGIF_ -> SWLRSSS (in isoform 4).
FT                                  /FTId=VSP_004275.
FT   CONFLICT     188     188       R -> K (IN REF. 3).
FT   CONFLICT     220     220       Y -> H (IN REF. 3).
FT   CONFLICT     276     276       T -> A (IN REF. 3).
FT   CONFLICT     584     584       K -> E (IN REF. 3).
FT   CONFLICT     628     628       T -> A (IN REF. 3).
FT   CONFLICT     630     630       K -> E (IN REF. 3).
FT   CONFLICT     682     682       K -> N (IN REF. 3).
SQ   SEQUENCE    2230 AA;  261139 MW;  3BB733DB1EA86134 CRC64;
```

```
Query Match            3.7%;  Score 197;  DB 1;  Length 2230;
Best Local Similarity 17.9%;  Pred. No. 0.058;
Matches 185;  Conservative 212;  Mismatches 366;  Indels 270;  Gaps  46;

Qy     15 ISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVELLISTLEKGVWHLG  74
             ::|::|      |  |
Db    384 IAQLRSRIKQ---------MTTQGEELREQKEKSERAA--FEELEKALSTAQK-------  425

Qy     75 WTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNLLQPTIVDKLLVRDV 134
                                 ||           : :|:|    |
Db    426 -TEE---------------ARRKLKAEM-DEQIKTIEKTSEE--ERISLQQELSRVKQEVVDV 469

Qy    135 LDKCMEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSAFLNVLRQTGNNELV 194
                        |      :  |          |:     :|
Db    470 MKKSSEEQIAKLQKLHEKELARKE---------QELTKKLQTREREFQEQMKVALEKSQSEYL 523

Qy    195 QELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMENNSSESSFADSSV 254
          :                       :  ||:      :     : :::
Db    524 K---ISQEKEQQESSLA-LEELELQKKAILT----------------ESENKLRDLQQ 560

Qy    255 VSESDTSLAEGSVSCLDESLGHNSNMGSDSGT-MGSDSDEENVAARASPE---PELQLRP 310
                                       |:  ::: :  ||              ||:
Db    561 EAETYRIRILELESSLEKSLQENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESLK 620

Qy    311 YQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNKVLL-- 368
          |       |                  |:                :   : |::
Db    621 HQ----QDAL----------WTEKLQVLKQQYQTEMEKLREKCEQEKETLLKDKEIIFQ 665

Qy    369 --VEQLFRKEFQPFLKKWYRYVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLEN 426
             ::  :        |:  : :: |: :   |      ::::       ||
Db    666 AHIEEMNEKTLEKLDVKQTELESLSSE-------LSEVLKA------RHKLEEELSVLKD 712

Qy    427 GEDAGVQLSDFSLIIIDECHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQILG 486
                                           :  :::|     |      ::|
Db    713 QTDKMKQELEAKM---DEQKNHHQQQVDSIIKEH----EVSIQRTEKALKDQINQLELL- 764

Qy    487 LTASPGVGGATKQAKABEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAI----A 542
                                     :  : : :|| : :::: :  |     |: :
Db    765 --------LKERDKHLKEHQAH-----VENLEADIKRSEGELQQASAKLDVFQSYQ 807

Qy    543 DATRED--PFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC 600
          ::       ::     :            :        :|::
Db    808 SATHEQTKAYEEQLAQLQQK----LLDLETE----RILLTKQVAEVEAQKKDVC 853

Qy    601 AE---HLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSEGGDDEYCDGD 657
                 :  :  ::: : :::    : ::|
Db    854 TELDAH-----KIQVQDLMQQLEK----QNSEMEQKVKSLTQVVESKLE------DGN 896

Qy    658 EDEDDLKKPLKLDETDRFLMTLFENNKMLKRLAE----------NPEYE--------- 697
```

```
Db    897 ::::|:|:::    |:     |: ::: :   |    |||
          KEQEQTKQILV--EKENMILQMREGQKKEIBILTQKLSAKEDSIHILNEEYETKFKNQEK 954

Qy    698 -----NEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAH 752
          ::|:         :|:|:       |||        ::
Db    955 KMEKVKQKAKEMQETLKKKLLDQEAKLK---KELENTALELSQ--KEKQFNAKMLEMA 1007

Qy    753 HLIGAGHSSEFKPM--TQNEQKEVISKFRTGKINLLIA--------TTVAEEGLDIKECNI- 803
          :       ::|::         |||     ||:
Db   1008 QANSAGISDAVSRLETNQKEQIESLTEVHRRELANDVISIWEKKLNQQAEELQEIHEIQLQ 1067

Qy    804 --------VIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIBHETVNDFRE 851
                  :                            |:         |:
Db   1068 BKEQEVAELKQKILLFGCEKKEM---------NKEITWL---KEEGVKQDTLNELQE 1113

Qy    852 KMMYKAIHCVQNMKPEE---YAH-KILE------LQMQSIMEKKMKTKRNIAKHYKNNPS 901
          :       :|  |:|          ||:         :||
Db   1114 QLKQKSAH-VNSLAQDETKLKAHLEKLEVDLNKSLKENTFLQEQLVELKMLAEEDKRKVS 1172

Qy    902 LITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQ----- 955
          :|:         ::                |:|:  :|:      :|:::|:
Db   1173 ELT-----SKLKTTDEEFQSLKSSHE-------------KSNKSLEDKSLEFKKLSEEL 1213

Qy    956 -INGEIICKCGQA 967
          :   |:
Db   1214 AIQLDICCKKTEA 1226
```

Search completed: June 2, 2004, 19:10:47
Job time : 24 secs

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:       June  2, 2004, 19:06:53 ; Search time 49 Seconds
                                         (without alignments)
                                         6600.126 Million cell updates/sec

Title:        US-09-515-363C-2
Perfect score: 5311
Sequence:     1 MSNGYSTDENFRYLISCFRA.........LPITFPNLDYSECCLFSDED 1025

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5

Searched:     1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:     1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    SPTREMBL_25:*
              1: sp_archea:*
              2: sp_bacteria:*
              3: sp_fungi:*
              4: sp_human:*
              5: sp_invertebrate:*
              6: sp_mammal:*
              7: sp_mhc:*
              8: sp_organelle:*
              9: sp_phage:*
             10: sp_plant:*
             11: sp_rodent:*
             12: sp_virus:*
             13: sp_vertebrate:*
             14: sp_unclassified:*
             15: sp_rvirus:*
             16: sp_bacteriap:*
             17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	5311	100.0	1025	4	Q9H3G6	Q9H3G6	homo sapien
2	5285	99.5	1025	4	Q9BYX4	Q9BYX4	homo sapien
3	4199	79.1	1025	11	Q8R5F7	Q8R5F7	mus musculu
4	4194	79.0	1025	11	Q8K5C7	Q8K5C7	mus musculu
5	2673	50.3	693	11	Q8BZ01	Q8BZ01	mus musculu
6	2633	49.6	683	11	Q8BYC9	Q8BYC9	mus musculu
7	2456	46.2	468	4	Q96MX8	Q96MX8	homo sapien
8	2271.5	42.8	514	11	Q99KS4	Q99KS4	mus musculu
9	2059.5	38.8	467	11	Q8R144	Q8R144	mus musculu
10	1356	25.5	678	11	Q99J87	Q99J87	mus musculu
11	1352	25.5	678	4	Q96C10	Q96C10	homo sapien
12	1344	25.3	678	4	Q9HAM6	Q9HAM6	homo sapien
13	1122	21.1	244	11	Q9D2Z5	Q9D2Z5	mus musculu
14	1086.5	20.5	925	4	O95786	O95786	homo sapien
15	1071	20.2	221	4	Q86X56	Q86X56	homo sapien
16	1018	19.2	940	6	Q9GLV6	Q9GLV6	sus scrofa
17	687	12.9		5	O44165	O44165	caenorhabdi
18	678	12.8		11	Q8VE79	Q8VE79	mus musculu
19	555.5	10.5		11	Q8C7T2	Q8C7T2	mus musculu
20	555	10.5	1164	11	Q93413	Q93413	caenorhabdi
21	516	9.7	398	4	Q9NT04	Q9NT04	homo sapien
22	516	9.7	620	5	Q17545	Q17545	caenorhabdi
23	453	8.5	752	17	Q9V1Z5	Q9V1Z5	pyrococcus
24	441.5	8.3	650	17	O59524	O59524	pyrococcus
25	431.5	8.1	1883	10	Q8LMR2	Q8LMR2	oryza sativ
26	417	7.9	741	17	O28814	O28814	archaeoglob
27	395.5	7.4	764	17	Q8TZH8	Q8TZH8	pyrococcus
28	391	7.4	410	17	Q8C513	Q8C513	mus musculu
29	389	7.3	821	17	Q8TUH1	Q8TUH1	methanosarc
30	384	7.2	1604	10	Q7XQ14	Q7XQ14	oryza sativ
31	368	6.9	864	17	Q8PX35	Q8PX35	methanosarc
32	365.5	6.9	837	5	Q86L44	Q86L44	dictyosteli
33	350	6.6	513	17	Q978A0	Q978A0	thermoplasm
34	343	6.5	182	11	Q9D1X4	Q9D1X4	mus musculu
35	343	6.5	2042	10	Q9M9P8	Q9M9P8	arabidopsis
36	342	6.4	738	17	O27466	O27466	methanobact
37	332.5	6.3	1383	5	Q9VDA0	Q9VDA0	drosophila
38	332	6.3	508	17	Q9HI46	Q9HI46	thermoplasm
39	313.5	5.9	784	17	Q9HMW5	Q9HMW5	halobacteri
40	309.5	5.8	1883	5	Q95ZG7	Q95ZG7	dictyosteli
41	302.5	5.7	741	17	Q8TUS6	Q8TUS6	methanopyru
42	302.5	5.7	1789	5	Q8T145	Q8T145	dictyosteli
43	295.5	5.6	783	3	Q9HE09	Q9HE09	schizosacch
44	295	5.6	1458	10	Q8W367	Q8W367	oryza sativ
45	295	5.6	1458	10	Q7XD96	Q7XD96	oryza sativ

ALIGNMENTS

```
RESULT 1
Q9H3G6
ID   Q9H3G6          PRELIMINARY;          PRT;    1025 AA.
AC   Q9H3G6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Melanoma differentiation associated protein-5.
GN   MDA5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RA   Kang D.-C., Fisher P.B.;
RT   "Melanoma differentiation associated gene-5 (mda-5), an interferon
RT   inducible gene of limited homology to RNA helicase.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF095844; AAG34368.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE  1025 AA;  116687 MW;  4BBE75491D863741 CRC64;

Query Match             100.0%;  Score 5311;  DB 4;  Length 1025;
Best Local Similarity   100.0%;  Pred. No. 2.7e-288;
Matches 1025;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIRTVATSGNMQAVE 60
```

```
Q9BYX4
ID   Q9BYX4              PRELIMINARY;         PRT;   1025 AA.
AC   Q9BYX4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA helicase-DEAD box protein RH116.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cocude C., Kolesnitchenko V., Billaut-Mulot O., Truong M.-J.,
RA   Capron A., Bahr G.M.;
RT   "Identification of a new RNA helicase (RH116) regulated by the
RT   immunomodulator Murabutide.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY017378; AAG54076.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   1025 AA;  116671 MW;  6B4F3D14E88A7D6D CRC64;

Query Match             99.5%;  Score 5285;  DB 4;  Length 1025;
Best Local Similarity   99.5%;  Pred. No. 7.7e-287;
Matches 1020;  Conservative   2;  Mismatches   2;  Indels   3;  Gaps   0;
```

```
Qy      1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE   60

Qy     61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL  120

Qy    121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA  180

Qy    181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM  240

Qy    241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARA  300

Qy    301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 SPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVI  360

Qy    361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQIIENS  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 VLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQIIENS  420

Qy    421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIP  480

Qy    481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA  540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 LPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFA  540

Qy    541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 IADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVC  600

Db    601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE  660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 AEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDE  660

Db    661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR  720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 DDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESAR  720

Db    721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT  780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRT  780

Db    781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV  840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 GKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGV  840

Db    841 IEHETVNDFREKMMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNP  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 IEHETVNDFRERMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNLAKHYKNNP  900

Db    901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI  960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 SLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEI  960

Db    961 ICKCCGQAWGTMVHKGLDLPCLKIRNFVVVFKQNSTKKQYKKWVELPITFPNLDYSECCL 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 ICKCCGQAWGTMVHKGLDLPCLKIRNFVVVFKQNSTKKQYKKWVELPITFPNLDYSECCL 1020

Db   1021 FSDED 1025
          |||||
Qy   1021 FSDED 1025
```

RESULT 2

```
Qy     1  MSNGYSTDENFRYLLSCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE   60
          :::.: ::: .:.:::.:.: :::::::::::::.: :: :::::: .::: :: ::::: 
Db     1  MSIVCSAEDSFRNLLFFFRPRLKMYIQVEPVLDHLIFLSAETKEQILKKINTCGNTSAAE   60

Qy    61  LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL  120
          ||||||||.|.|.|||..||||.|.|.|.|||||.|.||||||||||..|||||||.||.
Db    61  LLLSTLEQGQWPLGWTQMFVEALEHSGNPLAARYVKPTLTDLPSPSSETAHDECLHLLTL  120

Qy   121  LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA  180
          ||||||||||..|:||.|.|..|||.|||||||:||.:||||||||||||||||||||||
Db   121  LQPTLVDKLLINDVLDTCFEKGLLTVEDRNRISAAGNSGNESGVRELLRRIVQKENWFST  180

Qy   181  FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM  240
          ||:||||||||.|.|||||..|.|.:..:.||:..:.||..||.||.||::|::||||:|
Db   181  FLDVLRQTGNDALFQELTGGGCPEDNTDLANSSHRDGPAANECLLPAVDESSLETEAWNV  240

Qy   241  ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENV-AAR  299
          ::..: . .:|||||.||||||||||||| .|||||||||.|||||||||||||...:.:
Db   241  DDILPEASCTDSSVVTESDTSLAEGSVSCPEDSLGHNSNMGRDSGTMGSDSDESVIQTKR  300

Qy   300  ASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKV  359
          .|||||||||||||||||||||.|||||||||||||||||||:.||||||||.|||.|||
Db   301  VSPEPELQLRPYQMEVAQPALDGKNIIICLPTGSGKTRVAVYITKDHLDKKKQASESGKV  360

Qy   360  IVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILEN  419
          ||||||||:.||||||||.|.||||||:|||||||||||||||||||.|:|||||||||||
Db   361  IVLVNKVMLAEQLFRKEFNPYLKKWYRIIGLSGDTQLKISFPEVVKSYDVIISTAQILEN  420

Qy   420  SLINLENGEDAGVQLSDFSLIIIDECHHTNKEAYNNIMRHYLMQKLKNNRLKKENKPVI   479
          ||||||.|.|..||||||||||||||||||||||:||||:|||.||:||.|||:|||:|
Db   421  SLINLESGDDDGVQLSDFSLIIIDECHHTNKEAVVNNIMRRYLKQKLRNNDLKKQNKPAI  480

Qy   480  PLPQILGLTASPGVGGATKQAKABEHILKLCAN.DAFTIKTVKENLDQLKNQIQEPCKKF  539
          |||||||||||||||.|.|||:|.:|||.:|||.|||||||||||:.||:|||:|||||||
Db   481  PLPQILGLTASPGVGAAKKQSEABKHILNICANLDAFTIKTVKENLGQLHQIKEPCKKF   540

Qy   540  AIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERV  599
          .|||.||..|||||||||||.|||||||.|||||||||.|||||||||||||..||||.|
Db   541  VIADDTRENPFKEKLLEIMASIQTYCQKSPMSDFGTQHYEQWAIQMEKKAAKDGNRKDRV  600

Qy   600  CAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSEGGDDEYCDGDED   659
          |||||||||:|||||||||||||||||||||.:.|:|||||.:.|.:..||::..|:.|.
Db   601  CAEHLRKYNKALQINDTIRMIDAYTHLETFYTDEKEKKFAVL-NDSDKSDDEASSCNDQL  659

Qy   660  EDDLKKPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESA  719
          .:|:||.|||||||.||||||||:|||||:||||||:|||||||||||||||.:||:|||.
Db   660  KGDVKKSLKLDETDEFLMNLFFDNKKMLKKLAENPKYENEKLIKLRNTILEQFTRSEESS  719

Qy   720  GIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFR   779
          |||||||||||.||||||.||.||||||||||||||||||||.||||||.||||||||||
Db   720  RGIIFTKTRQSTYALSQWIMENAKFAEVGVKAHHLIGAGHSSEVKPMTQTEQKEVISKFR  779

Qy   780  TGKINLLIATTVABEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSG  839
          ||.|||||||||||||||||||||||||||||||||||||||||||||||||||.:|||||
Db   780  TGHINLLIATTVABEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVTSSGSG  839

Qy   840  VIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNN  899
          |.|.|.||||||||||||||..|||||||||||||||||.||:||||||.|||||.:|:|
Db   840  VTEREIVNDFREKMMYKAINRVQNMKPEEYAHKILELQVQSILEKKMKVKRSIAKQYNDN  899

Qy   900  PSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGE  959
          ||||||||||||||.||||.|||||||||||||||||:|||||||||||||.|||||.||
Db   900  PSLITLLCKNCSMLVCSGENIHVIEKMHHVNMTPEFKGLYIVRENKALQKKFADYQTNGE  959

Qy   960  IICKCGQAWGTMVHKGLDLPCLKIRNFVVVFQNSTKKQYKKWVELPITFPNLDYSECC  1019
          |||||||||||||||||||||||||||||||:..||||||||||||||.||:|||.||.
Db   960  IICKCGQAWGTMVHKGLDLPCLKIRNFVVNFKNNSPKKQYKKWVELPIRFPDLDYSEYC  1019

Qy  1020  LFSDED  1025
          |:||||
Db  1020  LYSDED  1025
```

RESULT 3

Q8R5F7

```
ID   Q8R5F7       PRELIMINARY;       PRT;   1025 AA.
AC   Q8R5F7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   HELICARD.
GN   MDA5 OR 9130009C22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kovacsovics M., Hofmann K., Tschopp J.;
RT   "HELICARD, a novel CARD-containing helicase.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY077205.1; AAL77205.1; -.
DR   MGD; MGI:1918836; Mda5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE  1025 AA;  115971 MW;  708FCAC690C9F6D8 CRC64;
```

Query Match   79.1%; Score 4199; DB 11; Length 1025;
Best Local Similarity 79.7%; Pred. No. 3.6e-226;
Matches 818; Conservative 75; Mismatches 131; Indels 2; Gaps 2;

```
RESULT 4
Q8K5C7
ID   Q8K5C7;           PRELIMINARY;      PRT;   1025 AA.
AC   Q8K5C7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Melanoma differentiation associated gene 5-like protein.
GN   MDA5 OR 9130009C22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Kang D.-C., Fisher P.B.;
RT   "Murine homolog of melanoma differentiation associated gene-5 - a DEXH
RT   group RNA dependent ATPase.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF374384; AAM21359.1; -.
DR   MGD; MGI:1918836; Mda5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE 1025 AA; 115941 MW; 8338A8510C5ED3A8 CRC64;

Query Match             79.0%;  Score 4194;  DB 11;  Length 1025;
Best Local Similarity   79.6%;  Pred. No. 6.9e-226;
Matches  817;  Conservative  75;  Mismatches  132;  Indels  2;  Gaps  2;

Qy       1 MSNGYSTDENFRYLLSCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
            ||||||  :::||| :|||  ||||||||||| | |||| ||||||| ||| | |||
Db       1 MSIVCSAEDSFRNLLLFFPRPRLKMYIQVEPVLDLHILFLSAETKEQILKKINTCGNTSAAE 60

Qy      61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
            |||||||:|||||||||||||||||| |||  ||| |||| |||||||| || |||||
Db      61 LLLSTLEQGQWPLGWTQMFVEALEHSGNPLAARYVKPTLTDLPSPSSETAHDECLHLLTL 120

Qy     121 LQPTIVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
            |||||||||||  :|  ||  ||| |:||||||| |||||||||||||||||||||||
Db     121 LQPTIVDKLLINDVLDTCFEKGLLTVEDRNRISAAGNSGNESGVRELLRRIIVQKENWFST 180

Qy     181 FLNVLRQTGNNELVQELTGSDCSRSNAEIENLSQVDGPQVEEQLLSTTVQPNLKEVWGM 240
            ||| ||||| |  :|:||:|| |: ||  | ||| | |   ||  :|:|   ::| |||
Db     181 FLDVLRQTGNDALFQELTGGCPEDNTDLANSSHRDGPAANECLLPAVDESSLETEAWNV 240

Qy     241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENV-AAR 299
            :: ||| :|:: :: | :||:| |||| || :|  ||:| | ||| |||||:  |||
Db     241 DDILPEASCTDSSVTTESDTSLAEGSVSCFDESLGHNSNMGRDSGTMGSDSDESVIQTKR 300

Qy     300 ASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKV 359
            ||||||||||||||||||||| |||||||||||||||||||| ||||||||| |||:||
Db     301 VSPEPELQLRPYQMEVAQPALDGKNIIICLPTGSGKTRVAVYITKDHLDKKKQASESGKV 360

Qy     360 IVLVNKVLLIVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILEN 419
            ||||||:||| :||||:|| || |||||:||||||||||||||||||:||:||||||||||
Db     361 IVLVNKVMLAEQLFRKEFNPFLKKWYRIIGLSGDTQLKISFPEVVKSYDVIISTAQILEN 420

Qy     420 SLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVI 479
            |||||| ||:||  | |||||| |||||||||||||||||||||||:|||:|||||||
Db     421 SLLNLESGDDDGVQLSDFSLIIIDECHHTNKEAVYNNIMRRYLKQKLRRNDLKKQNKPAI 480

Qy     480 PLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKF 539
            |||||||||||||||| || ||||||||| |||||||||||||||| |||:| | |||||
```

```
Db     481 PLPQILGLTASPGVGAAKKQSEAEKHILNICANLDAFTIKTVKENLGQLKHQIKEPCKKF 540
            ||||||||||||||  || || ||||||| |||||||||||||||| |||:| ||||||
Qy     540 AIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERV 599
            :|:|::|| ||||||| |: | |||| | |||||| || ||||| |||||  || :|:|
Db     541 VIADDTRENPFKEKLLEIMASIQTYCQKSPMSDFGTQHYEQWAIQMEKKAAKDGNRKDRV 600
Qy     600 CAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSEGGDEYCGDGDED 659
            ||||||||||||||||||||||||| :||  ::|  ::|||| | :: :| :  :
Db     601 CAEHLRKYNEALQINDTIRMIDAYSHLEAFYTDEKEKKFAVL-NDSDKSDDBASSCNDQL 659
Qy     660 EDDLKKPLKLDETDRFLMTLFFENNKMLKRIAENPEYENEKLTKLRNTIMEQYTRTEESA 719
            ::||| |  ||||: |:||| |  :| |:: |||| |||||| | ||||:: ::  :|||
Db     660 KGDVKKSLKLDETDEFIMNLFFDNKKMLKKLAENPKYENEKLIKLRNTILEQFTRSEESS 719
Qy     720 RGIIFKTKRQSAYALSQWITTENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFR 779
            |||||||||||||||||||| ||||||||||||||||||||||||  ||:|| ||||||||
Db     720 RGIIFKTKRQSTYALSQWIMENAKFAEVGVKAHHLIGAGHSSEVKPMTQTEQKEVISKFR 779
Qy     780 TGKINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHGSGS 839
            |||||||||||||||||||: |||||||||||||||||||||||||||||||| | ||||
Db     780 TGEINLLIATTVAEEGLDIK3CNIVIRYGLVTNEIAMVQARGRARADESTYVLVTSSGSG 839
Qy     840 VIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKMKTKRNIAKHYKNN 899
            |  ||||||||||||||| ||: :||||||||||||||:|| :::|: ||||: |:|:
Db     840 VTEREIVNDFREKMMYKAINRVQNMKPEEYAHKILELQVQSILEKKMKVKRSIAKQYNDN 899
Qy     900 PSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGE 959
            ||||||||||||:|||||||||||||||||||||||||||||||||||||  ||| |||
Db     900 PSLITLLCKNCSMLVCSGENIHVIEKMHHVNMTPEFKGLYIVRENKALQKKFADYQTNGE 959
Qy     960 IICKCGQAWGTMVHKGLDLPCLKIRNFVVVFKRNSTKKQYKKWVELPITFPNLDYSECC 1019
            |||||||||||||||||||||||||||||  ||||||| ||||||||||| |  |||||
Db     960 IICKCGQAWGTMVHKGLDLPCLKIRNFVVNFKRNNSPKKQYKKWVELPIRFPDLDYSEYC 1019
Qy    1020 LFSDED 1025
            | ||||
Db    1020 LYSDED 1025
```

```
RESULT 5
Q8BZO1
ID   Q8BZO1;           PRELIMINARY;      PRT;   693 AA.
AC   Q8BZO1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to melanoma differentiation associated protein-5
DE   (Fragment).
GN   MDA5 OR 9130009C22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Vagina;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   The RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK037057; BAC29687.1; -.
DR   MGD; MGI:1918836; Mda5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   InterPro; IPR001410; DEAD.
DR   Pfam; PF00270; DEAD; 1.
DR   SMART; SM00487; DEXDc; 1.
FT   NON_TER   693   693
SQ   SEQUENCE 693 AA; 77757 MW; BB11B45BB5BE019D CRC64;

Query Match             50.3%;  Score 2673;  DB 11;  Length 693;
```

Left column

Best Local Similarity 75.6%; Pred. No. 3.8e-141;
Matches 525; Conservative 60; Mismatches 107; Indels 2; Gaps 2;

```
Qy    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
          |  :::||| ||||||||||||||| ||:||::||: |::
Db    1 MSIVCSAEDSFRNLILFFRPRLKMYIQVEPVLDHLIFLSAETKEQILKKINTCGNTSAAE 60

Qy   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
          ||||||||| |:|
Db   61 LLLSTLEQGQWPLGWTQMFVEALEHSGNPLAARYVKPTLTDLPSPSSETAHDECLHLLTL 120

Qy  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
          |||||||||      ||||||:||||||||||||:|:||||||||||||||||||||||
Db  121 LQPTLVDKLLINDVLDTCFEKGLLTVEDRNRISAAGNSGNESGVRELLRRIVQKENWFST 180

Qy  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
          |:|| |||||||   ::||||||||
Db  181 FLDVLRQTGNDALFQELTGGGCPEDNTDLANSSHRDGPAANECLLPAVDESSLETEAWNV 240

Qy  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENV-AAR 299
          :        |||||||||:||:||||
Db  241 DDILPRASCTDSSVTTESDTSLAEGSVSCFDESLGHNSNMGRDSGTMGSDSDESVIQTKR 300

Qy  300 ASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKV 359
                |:||||||||||||||||||||||||||||||:  ::
Db  301 VSPEPEIQLRPYQMEVAQPALDGKNIIICLPTGSGKTRVAVYITKDHLDKKKQASESGKV 360

Qy  360 IVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILEN 419
          ||||||:|:||||||:|||||||||||||||||||||||||||
Db  361 IVLVNKVMLAEQLFRKEFNPYLKKWYRIIGLSGDTQLKISFPEVVKSYDVIISTAQILEN 420

Qy  420 SLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVI 479
          |||||| |   | ||||||||||||||:||||||:| |||:||||||| :||||||||
Db  421 SLLNLESGDDDGVQLSDFSLIIIDECHHTNKEAVYNNIMRRYLKQLRNNDLKKQNKPAI 480

Qy  480 PLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKF 539
          |||||||||||||||  |||  ||::|| |||:|||||||||||||| :|||:|||||||
Db  481 PLPQILGLTASPGVGAAKKQSEAEKHILNICANLDAFTIKTVKENLGQLKHQIKEPCKKF 540

Qy  540 AIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERV 599
          :||| |||:||||||:|||| ||||||||||||||:|||||||||||||||||| ||:|
Db  541 VIADDTRENPFKEKLLEIMASIQTYCQKSPMSDFGTQHYEQWAIQMEKKAAKDGNRKDRV 600

Qy  600 CAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKKFAVIEDDSDEGGDDEYCDGDED 659
          |||||||||||||||||||||||||:||||||  |:|| |||::||       ::|||:|
Db  601 CAEHLRKYNEALQINDTIRMIDAYSHLETFYTDEKEKKFAVL-NDSDKSDDEASSCNDQL 659

Qy  660 EDDLKKPLKLDETDRFLMTLFFENNKMLKRLAEN 693
          : :|| ||||||||||| :|:
Db  660 KGDVKKSLKLDETDEFLMNLFFDN 683
```

RESULT 6
Q8BYC9
ID   Q8BYC9;
AC   Q8BYC9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to melanoma differentiation associated protein-5
DE   (Fragment).
GN   MDA5 OR 9130009C22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of

Right column

RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK040519; BAC30614.1; -.
DR   MGD; MGI:1918836; Mda5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   InterPro; IPR001410; DEAD.
DR   Pfam; PF00270; DEAD; 1.
DR   SMART; SM00487; DEXDc; 1.
FT   NON_TER   683   683
SQ   SEQUENCE   683 AA; 76572 MW; 012F0414B28251C3 CRC64;

Query Match   49.6%; Score 2633; DB 11; Length 683;
Best Local Similarity 75.6%; Pred. No. 6.4e-139;
Matches 517; Conservative 59; Mismatches 106; Indels 2; Gaps 2;

```
Qy    1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
          |  :::||| ||||||||||||||| ||:||::||: |::
Db    1 MSIVCSAEDSFRNLILFFRPRLKMYIQVEPVLDHLIFLSAETKEQILKKINTCGNTSAAE 60

Qy   61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
          ||||||||| |:|
Db   61 LLLSTLEQGQWPLGWTQMFVEALEHSGNPLAARYVKPTLTDLPSPSSETAHDECLHLLTL 120

Qy  121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQKENWFSA 180
          |||||||||      ||||||:||||||||||||:|:||||||||||||||||||||||
Db  121 LQPTLVDKLLINDVLDTCFEKGLLTVEDRNRISAAGNSGNESGVRELLRRIVQKENWFST 180

Qy  181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGM 240
          |:|| |||||||   ::||||||||
Db  181 FLDVLRQTGNDALFQELTGGGCPEDNTDLANSSHRDGPAANECLLPAVDESSLETEAWNV 240

Qy  241 ENNSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENV-AAR 299
          :        |||||||||:||:||||
Db  241 DDILPRASCTDSSVTTESDTSLAEGSVSCFDESLGHNSNMGRDSGTMGSDSDESVIQTKR 300

Qy  300 ASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKV 359
                |:||||||||||||||||||||||||||||||:  ::
Db  301 VSPEPEIQLRPYQMEVAQPALDGKNIIICLPTGSGKTRVAVYITKDHLDKKKQASESGKV 360

Qy  360 IVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILEN 419
          ||||||:|:||||||:|||||||||||||||||||||||||||
Db  361 IVLVNKVMLAEQLFRKEFNPYLKKWYRIIGLSGDTQLKISFPEVVKSYDVIISTAQILEN 420

Qy  420 SLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVI 479
          |||||| |   | ||||||||||||||:||||||:| |||:||||||| :||||||||
Db  421 SLLNLESGDDDGVQLSDFSLIIIDECHHTNKEAVYNNIMRRYLKQLRNNDLKKQNKPAI 480

Qy  480 PLPQILGLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKF 539
          |||||||||||||||  |||  ||::|| |||:|||||||||||||| :|||:|||||||
Db  481 PLPQILGLTASPGVGAAKKQSEAEKHILNICANLDAFTIKTVKENLGQLKHQIKEPCKKF 540

Qy  540 AIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERV 599
          :||| |||:||||||:|||| ||||||||||||||:|||||||||||||||||| ||:|
Db  541 VIADDTRENPFKEKLLEIMASIQTYCQKSPMSDFGTQHYEQWAIQMEKKAAKDGNRKDRV 600

Qy  600 CAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKKFAVIEDDSDEGGDDEYCDGDED 659
          |||||||||||||||||||||||||:||||||  |:|| |||::||       ::|||:|
Db  601 CAEHLRKYNEALQINDTIRMIDAYSHLETFYTDEKEKKFAVL-NDSDKSDDEASSCNDQL 659

Qy  660 EDDLKKPLKLDETDRFLMTLFFEN 683
          : :|| ||||||||||| :|:
Db  660 KGDVKKSLKLDETDEFLMNLFFDN 683
```

RESULT 7
Q96MX8
ID   Q96MX8;   PRELIMINARY;   PRT;   468 AA.
AC   Q96MX8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ31731.

OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA   Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano Y.,
RA   Watanabe M., Fujimori K., Tanai H., Ishida H., Yamashita H., Chiba Y.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA   Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK056293; BAB71141.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   468 AA; 54470 MW; 33D4D055C7436DFA CRC64;

Query Match   46.2%; Score 2456; DB 4; Length 468;
Best Local Similarity   99.1%; Pred. No. 3e-129;
Matches 464; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   558 MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTI 617
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTI 60

Qy   618 RMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM 677
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 RMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM 120

Qy   678 TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW 737
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW 180

Qy   738 ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLD 797
             |||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db   181 ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGRINLLIATTVAEEGLD 240

Qy   798 IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKA 857
             ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db   241 IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIERETVNDFREKMMYKA 300

Qy   858 IHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG 917
             |||||||||||||||||||:||||:||||||:|:|||:|::|||||:|||||||||||||
Db   301 IHCVQNMKPEEYAHKILELQVQSILEKKMKVKSIAKQYNDNPSLITLLCKNCSVLACSG 360

Qy   918 EDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMMVHKGL 977
             ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db   361 EDIHVIEKMHHVNMTPEFKELYIVRENKTLQKKCADYQINGEIICKCGQAWGTMMVHKGL 420

Qy   978 DLPCLKIRNFVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
             |||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DLPCLKIRNFVVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 468

RESULT 8
Q99KS4
ID   Q99KS4;   PRELIMINARY;   PRT; 514 AA.
AC   Q99KS4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).

GN   MDA5 OR 9130009C22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC004031; AAH04031.1; -.
DR   MGD; MGI:1918836; Mda5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
FT   NON_TER   1   1
SQ   SEQUENCE   514 AA; 59429 MW; 9025A02BE6D80F1F CRC64;

Query Match   42.8%; Score 2271.5; DB 11; Length 514;
Best Local Similarity   85.0%; Pred. No. 6.9e-119;
Matches 438; Conservative 30; Mismatches 46; Indels   1; Gaps   1;

Qy   511 ANLDAFTIKTVKENLDQLKNQIQBPCKKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPM 570
             |||||||||||||||:|||::||:|||||||:||| :|::|||||||||||::||||||:|||
Db   1 ANLDAFTIKTVKENLGQLKHQIKEPCKKKFVIADDTRENPFKEKLLEIMASIQTYCQKSPM 60

Qy   571 SDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFY 630
             ||||||:||||||||||||||:|||||:|||||||||||||||||||||||||:||||||
Db   61 SDFGTQHYEQWAIQMEKKAAKDGNRKDRVCAEHLRKYNEALQINDTIRMIDAYSHLETFY 120

Qy   631 NEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRL 690
             :||::|||||| :|||| : :||||||:|::||||:|||||||:|||.:||:.:||||:|
Db   121 TDEKEKKFAVL-NDSDESDDEASSCNDQLKGDVKKSLKLDETDEFLMNLFFDNKKMLKKL 179

Qy   691 AENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVK 750
             ||||:|||||:|||||||:|||:||:|||:||||||||||:|||||||:||:|||||||||
Db   180 AENPKYENEKLIKLRNTILEQFTRSEESSRGIIFTKTRQSTYALSQWIMENAKFAEVGVK 239

Qy   751 AHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLDIKECNIVIRYGLV 810
             ||||||||||||:||||||:||||||||||||:||||||||||||||||||||||||||||
Db   240 AHHLIGAGHSSEVKPMTQTEQKEVISKFRTGEINLLIATTVAEEGLDIKECNIVIRYGLV 299

Qy   811 TNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPEEYA 870
             |||||||||||||||||||||||.:||||||||:.||||||||||||||:.|||||||||||
Db   300 TNEIAMVQARGRARADESTYVLVTSSGSGVTERBIVNDFREKMMYKAINRVQNMKPEEYA 359

Qy   871 HKILELQMQSIMEKKMKTKENIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVN 930
             |||||||:||||:|||||:|:|||:|:::|||||:|||||||:|.|||::||||||||||
Db   360 HKILELQVQSILEKKMKVKSIAKQYNDNPSLITLLCKNCSMLVCSGENIHVIEKMHHVN 419

Qy   931 MTPEFKELYIVRENKALQKKCADYQINGEIICKCGQAWGTMMVHKGLDLPCLKIRNFVVV 990
             ||||||:|||||||||||||:||||:|||||||||||||||||||||||||||||||||||:
Db   420 MTPEFKGLYIVRENKALQKKFADYQTNGEIICKCGQAWGTMMVHKGLDLPCLKIRNFVVN 479

Qy   991 FKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED 1025
             |||||:||||||||||||:||:|:|||| |||:||
Db   480 FKNNSPKKQYKKWVELPIRFPDLLDYSEYCLYSDED 514

RESULT 9
Q8R144
ID   Q8R144;   PRELIMINARY;   PRT; 467 AA.
AC   Q8R144;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).

OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
EMBL; BC025508; AAH25508.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE 467 AA; 53966 MW; F3B0D976778F0442 CRC64;

Query Match   38.8%; Score 2059.5; DB 11; Length 467;
Best Local Similarity   84.6%; Pred. No. 4.2e-107;
Matches 396; Conservative 28; Mismatches 43; Indels 1; Gaps 1;

Qy   558 MTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHLRKYNEALQINDTI 617
               |||||||||||||||| ::|||||| ::||||||||||||||||||||||||||||
Db   1 MASIQTYCQKSPMSDFGTQHYEQWAIQMEKKAAKDGNRKAAKDNRKDRVCAEHLRKYNEALQINDTI   60

Qy   618 RMIDAYTHLETFYNEEKDKFAVIEDDSDEGGDDEYCDGDEDEDDLKKPLKLDETDRFLM 677
               ||::||.||||||:||:||:||. |.:|:| :|:| : :|||||||||||||||||
Db   61 RMIDAHSHLETFYITDEKEKKFAVL-NDSDESDDEASSCNDQLKGDVKKSLKLDETDEFLM 119

Qy   678 TLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQW 737
               ||:||||||:||:||||||:||||||||||:||:||:||||||||||||||||||||
Db   120 NLFFDNKKMLKKGLAENPKYENEKLIKLRNTILEQFTRSEESSRGIIFTKTRQSTYALSQW 179

Qy   738 ITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIATTVAEEGLD 797
               |||||||||||||||||||||||||:||||||||:|||||||:||||||||||||||
Db   180 IMENAKFAEVGVKAHHLIGAGHSSEVKPMTQTEQKEVISKFRTGEINLLIATTVAEEGLD 239

Qy   798 IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKA 857
               ||||||||||||||||||||||||||||||||||||||||||||:|||:||||||||||
Db   240 IKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVTSSGSGVTEREIVNDFREKMMYKA 299

Qy   858 IHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLITFLCKNCSVLACSG 917
               ||:|||||||||||||:||||:|||||||:||:||||||||||||||||||||||:||
Db   300 INRVQNMKPEEYAHKILELQVQSILEKKMKVKRSIAKQYNDNPSLITLLCKNCSMLVCSG 359

Qy   918 EDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEIICKCCGQAWGTMMVHKGL 977
               |:||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db   360 ENIHVIEKMHHVNMTPEFKGLYIVRENKALQKKFADYQINGEIICKCCGQAWGTMMVHKGL 419

Qy   978 DLPCLKIRNFVVVFKNNSTKKQYKKKWVELPITFPNLDYSECCLFSDED 1025
               ||||||||||||:|||||:|||||||||||||:|||||||||:||||||
Db   420 DLPCLKIRNFVVNFKNNSPKKQYKKKWVELPIRFPDLDYSEYCLYSDED 467

RESULT 10
Q99J87
ID   Q99J87;   PRELIMINARY;   PRT;   678 AA.
AC   Q99J87;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (DNA segment, Chr 11, Lothar Hennighausen 2,
DE   expressed).
GN   D11LGP2E OR D11LGP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=11161808;

RA   Miyoshi K., Cui Y., Riedlinger G., Lehoczky J., Zon L., Oka T.,
RA   Dewar K., Hennighausen L.;
RT   "Structure of the Mouse Stat 3/5 Locus: Evolution from Drosophila to
RT   Zebrafish to Mouse.";
RL   Genomics 71:150-155(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF317000; AAK15475.1; -.
DR   EMBL; AF316999; AAK15474.1; -.
DR   EMBL; BC029209; AAH29209.1; -.
MGD; MGI:1931560; D11Lgp2e.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE 678 AA; 76726 MW; DC42B75A3AD376A8 CRC64;

Query Match   25.5%; Score 1356; DB 11; Length 678;
Best Local Similarity   41.7%; Pred. No. 1.4e-67;
Matches 299; Conservative 129; Mismatches 237; Indels 52; Gaps 12;

Qy   306 LQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKASEPGKVIVLVNK 365
               ::||||||:|||||||||||| ||:|||||||||:||||: .: |||::|:|:||||
Db   1 MELRPYQWEVILPALEGKNIIIWLPTGAGKTRAAAFVAKRHLE----TVDRGKVVVLVNR   56

Qy   366 VLLVEQLFRKEFQPPLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
               | ||||||||:|| |||||:||||||||||||:||||||:|||:|:|||||||:||:|:
Db   57 VHLVSQ-HAEEFRRMLDKHWTVTTLSGDMGSRAGFGLMARSHDLLICTAELLQLALNSSE 115

Qy   426 NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIL 485
               ::|||:|: ||:||||:|||||:|:|||:|||| ||::| : |||:| |:| |||| |
Db   116 --EDEHVELREFSLIIVDECHHTHKDTVNTILSRYLEQKLK----KAE----PLPQVL 164

Qy   486 GLTASPGVGGATKQAKAEBHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT 545
               ||||||||||||:|||:|||||:||||||:|||||:||.:|:||||||||||:||:|||
Db   165 GLTASPGTGGATKLQGAIDHILQLCANLDTCHIMSPKNCYSQLLMHNPKPCKQYDLCQRR 224

Qy   546 REDPFKEKLLEIMTRIQTYCQMSPM-SDFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL 604
               ||:|||||||||:|| :: :: : .||:|| |::|: | ::||::|| ||:||:| ||
Db   225 AQDPFGDLIKKLMNQIHQQLEMPDLKQQFGTQMYBQQVVQLCKDAAEAGLQEQRVYALHL 284

Qy   605 RKYNEALQINDTIRMDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLK 664
               |:|||:| |||| :| :|. ::|:||:: : :: :
Db   285 RRYNDALFIHDTVRARDALDMLQDFYDRERTTKTQMVRAES-------------- 325

Qy   665 KPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIF 724
               ||:|||||||| ::|::|:|:|: :: : : :::| : ||::| :: |:||:|:
Db   326 ------------WLLKLFDDHKNVLGQLAARGP-ENPKLEMLERILLKQF-GSPGHTRGIIF 373

Qy   725 TKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKIN 784
               ||||||::||:||||||||||||||||||||||||||:||| ||:||||||||||||:|
Db   374 TRTRQTASSLLLWLRQQPCLQTVGIKPQMLIGAGNTSQSTHMTQKDQQBVIQEFRDGILS 433

Qy   785 LLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHE 844
               |||:||||||||:|:||||||:|||:||||:||||:|:||||||:|:|: ||| |:|||
Db   434 LLVATSVAEEGLDIAQCNVVVRYGLLTNEISMVQARGRARAGQSVYSFLATEGSREMKRE 493

Qy   845 TVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLIT 904
               | :: |::|:: |:::|:|:|| :|: |||:||| | : :: |||:|||: :|| |
Db   494 LTNEALEVLMEKAVAVQKMDPDFEFKAKIRDLQQASLVKRAARAAHREIQQGQFLPEHVQ 553

Qy   905 FLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENK-ALQKKCADYQINGEIIC- 962
               ||| |:|| :| |::||:|||||:| |: |||:|::: ::: ||:|| |||:|||| |:
Db   554 LLCINCMVAVGYGSDLRKVEGTHVNVANFNFSVYTTSQNPVVINKVFKDWRPGGTIRCS 613

Qy   963 KCCGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKKWVELPITFPNLDYSECC 1019

```
Db  614 NCGEVWGFQMIYKSVTLPVLKIGS--MLLETPRGKIQAKKWSRVPFSIPVFDILQDC 668
            ||:|  ||::.|||| :|||||      ::  :  : :  ||||||| :|| |||||
```

RESULT 11

```
Q96C10
ID   Q96C10           PRELIMINARY;      PRT;   678 AA.
AC   Q96C10;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC014949; AAH14949.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   678 AA;  76612 MW;  859I7149C7313D06 CRC64;
```

Query Match   25.5%;   Score 1352; DB 4; Length 678;
Best Local Similarity 40.7%; Pred. No. 2.3e-67;
Matches 292; Conservative 136; Mismatches 237; Indels 52; Gaps 12;

```
Qy  306 LQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK 365
            ::||||||||   |::|  |.::.|||||||||:.:| .: :  |||:::    |||||||
Db    1 MELRSYQWEVIMPALEGKNIIIWLPTGAGKTRAAAYVAKRHLE---TVDGAKVVVLVNR 56

Qy  366 VLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
            ||||||||:|| ||:||| |||||||||:||||:||| ||||||| ::|||||||:||:|
Db   57 VHLVTQ-HGEEFRRMLDGRWTVTTLSGDMGPRAGFGHLARCHDLLICTAELLQMALTSPE 115

Qy  426 NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIL 485
            |||||||||||||:|||||||||| ||::|:|||||:|||| ||:||||||| |||||||
Db  116 --EEEHVELTVFSLIIVDECHHTHKDTVYNVIMSQYLELKLQRAQ-------PLPQVL 164

Qy  486 GLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT 545
            |||||||||||::|||||| |||.||||||| |:|: ||| || ||:| |||:|||||||
Db  165 GLTASPGTGGASKLDGAINHVLQLCANLDTWCIMSPQNCCPQLQEHSQQPCKQYNLCHRR 224

Qy  546 REDPFFEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL 604
            :||||:|||||: :  ::    :|:|  :|| |||  :||||||||||||::|::| |||
Db  225 SQDPFFGDLLKKLMDQIHDHLEMPELSRKFGTQMYEQQVVKLLSEAAALAGLQEQRVTALHL 284

Qy  605 RKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAVIEDDSDEGGDDEYCDGDEDEDDLK 664
            |::|::|: |::    :  .:. |  |||   ||  ||||:|  ||    |:-------C 322
Db  285 RRYNDALLIHDTVRAVDALAALAQDFYHREHVTKTQIL------------------

Qy  665 KPIKLDETDRFLMTLFPENNKMLKRLAENPEYENEKLTKLRNTIMEQYTRTEESARGIIF 724
            ||:||  |     |:|||| |:|:|      |:|:|  |:: |:: | |||  |||:||| 373
Db  323 ------AERRLLALFDDRKNELAHLATHGP-ENPKLEMLEKILQRQFS-SSNSPRGIIF

Qy  725 TKTRQSAYALSQWITENEKFABVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKPRTGKIN 784
            |:||:|| ||: :||:||:|| |:||  |||:||  :: |    |::|||   |::  ||  433
Db  374 TRTRQSAHSLLLWLQQQGLQTVDIRAQLLIGAGNSSQSTHMTQRDQQEVIQKFQDGTLN

Qy  785 LLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHE 844
            |:||| |||||||||||:|||:|||||||| ||||||||||:.::::|::|| |||| :
```

```
Db  434 LLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRARADQSVYAFVATEGSRELKRE 493
            |||||:||||||| ||||:|:|||||||||:||||||||||||.|::|:|| ||::|:|:
Qy  845 TVNDFREKMYKAIHCVQNMKPEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLIT 904
            ::||  |::|:|||||:::|:||  :|: | || ||:||| ||:|: ||  |  ||::
Db  494 LINEALETIMEQAVAAVQKMDQAEYQAKIRDLQQAALTKRAAQAQRENQRQQFPVEHVQ 553
            ||:|  |:::||:| |:|:::|  ||  ||:::|||   |:|:|::: |  ||| |||
Qy  905 FLCKNCSVLACSGEDIHVIEKMHIVNMTPEFKELY-IVRENKALQKKCADYQINGEICK 963
            ||||||||| |||:|||||::|| ::||:||:|||   ||:||||||| ||:||| |||
Db  554 LLCINCMVAVGHGSDLRKVEGTHVNVNPNFSNYYNVSRDPVVINKVFDWKPGGVISCR 613
            ||| ||:| ||| ||:|::||:|| ::|||| |:|: ||| :||  :| ||| |||| |
Qy  964 -CGQAWGTMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWELPITFPNLDYSECC 1019
             ||| |||| ||||||||||::|| |||||||||:||:||:||||||| |||:|:||
Db  614 NCGEVWGLQMIYKSVSKLPVLKVRS--MLLETPQGRIQAKKWSRVPFSVPDFDFLQHC 668
```

RESULT 12

```
Q9HAM6
ID   Q9HAM6           PRELIMINARY;      PRT;   678 AA.
AC   Q9HAM6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK021416; BAB13818.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   678 AA;  76642 MW;  0628A21A71856EE5 CRC64;
```

Query Match   25.3%;   Score 1344; DB 4; Length 678;
Best Local Similarity 40.6%; Pred. No. 6.6e-67;
Matches 291; Conservative 136; Mismatches 238; Indels 52; Gaps 12;

```
Qy  306 LQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNK 365
            ::||||||||   |::|  |.::.|||||||||:.:| .: :  |||:::    |||||||
Db    1 MELRSYQWEVIMPALEGKNIIIWLPTGAGKTRAAAYVAKRHLE---TVDGAKVVVLVNR 56

Qy  366 VLLVEQLFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQILENSLLNLE 425
            ||||||||:|| ||:||| |||||||||:||||:||| ||||||| ::|||||||:||:|
Db   57 VHLVTQ-HGEEFRRMLDGRWTVTTLSGDMGPRAGFGHLARCHDLLICTAELLQMALTSPE 115

Qy  426 NGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQIL 485
            |||||||||||||:|||||||||| ||::|:|||||:|||| ||:||||||| |||||||
Db  116 --EEEHVELTVFSLIIVDECHHTHKDTVYNVIMSQYLELKLQRAQ-------PLPQVL 164

Qy  486 GLTASPGVGGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADAT 545
            |||||||||||::|||||| |||.||||||| |:|: ||| || ||:| |||:|||||||
Db  165 GLTASPGTGGASKLDGAINHVLQLCANLDTWCIMSPQNCCPQLQEHSQQPCKQYNLCHRR 224

Qy  546 REDPFFEKLLEIMTRIQTYCQMSPMS-DFGTQPYEQWAIQMEKKAAKKGNRKERVCAEHL 604
            :||||:|||||: :  ::    :|:|  :|| |||  :||||||||||||::|::| |||
```

```
Db    225  SQDPFGDLLKKIMDQIHDHLEMPELSRKFGTQMYEQQVVKLSEAAALAGLQEQRVYALHL  284

Qy    605  RKYNEALQINDTIRMIDAYTHLETYNEEKDKKFAVIEDDSDEGGDEYCDGDEDEDDLK       664
           |:||:|| ||  ||::|| | |:: ||::
Db    285  RRYNDALLIHDTVRAVDALAALQDFYHREHVTKTQIL------------C--------     322

Qy    665  KPLKLDETDRFLMTLFFENNKMLKRLAENPEYENEKLITKLRNTIMEQYTRTEESARGIIF   724
           |: |: ||    |      :||| |  |||||:|    |:: |  ||||
Db    323  ------AERRLLALFDDRKNELAHLATHGP-ENPKLEMLEKILQRFS-SSNSPRGIIF      373

Qy    725  TKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVLSKFRTGKIN   784
           ||:|||| : | :     | | |||| ||:|  :|| :|  | |:|  ||:|:|  ||||
Db    374  TRTRQSAHSLLLWLQQQGLQTVDIRAQLLIGAGNSSQSTHMTQRDQQEVIQKFQDGTLN   433

Qy    785  LLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHE   844
           |:|||||||||||||| ||:|||||||||||||||||||||||||||||| |||||| :
Db    434  LLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRAWADQSVVAFVATEGSSRELKRE   493

Qy    845  TVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPSLIT   904
           ||||:|||||||| :| :|::||:|| |||| |||:|  :|   |:|   :||::
Db    494  LINEALETLMEQAVAAVQKMDQAEYQAKIRDLQQAALTKRAAQAAQRENQRQQFPVEHVQ   553

Qy    905  FLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELY-IVRENKALQKKCADYQINGEIICK   963
           |||:||| : | | :|||||| || ::|  |:|  ||| ||  |:|:|  :|:  ::|
Db    554  LLCINCMVAVGHGSDLRKVEGTHHVNVNPNFSNYYNVSRDPVVINKVFKDWKPGGVISCR   613

Qy    964  -CGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECC 1019
            |||||||||||||| :: : :|||  |:: :: ||:  :| |||| || ::|:||
Db    614  NCGEVWGLQMIYKSVKLPVLKVRS--MLLETPQGRIQAKKWKSRVFFSVPDFDFLQHC  668
```

RESULT 13
Q9D2Z5

```
ID   Q9D2Z5;          PRELIMINARY;    PRT;    244 AA.
AC   Q9D2Z5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Adult male cecum cDNA, RIKEN full-length enriched library,
DE   clone:9130009C22 product:similar to melanoma differentiation
DE   associated protein-5 (Fragment).
GN   MDA5 OR 9130009C22RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata D., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
```

```
RX   MEDLINE=21085660; PubMed=11217851;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690 (2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RX   MEDLINE=99279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RX   MEDLINE=20499374; PubMed=11042159;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RX   MEDLINE=20530911; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).; --
DR   EMBL; AK018602; BAB31303.2; -.
DR   MGD; MGI:1918836; Mda5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00271; helicase_C; 1.
FT   NON_TER       1      1
SQ   SEQUENCE  244 AA;  28154 MW;  4D742F304832948C CRC64;
```

```
Query Match             21.1%;  Score 1122;  DB 11;  Length 244;
Best Local Similarity   88.1%;  Pred. No.4.2e-55;
Matches  215;  Conservative  11;  Mismatches  18;  Indels  0;  Gaps  0;
```

```
Qy    782  KINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVI  841
           :||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  EINLLIATTVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVTSSGSGVT   60

Qy    842  EHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTKRNIAKHYKNNPS  901
           |:| |||||||||||||:||||||||||||||||||||||:::|||||||||||||||||
Db     61  EREIVNDFREKMMYKAINRVQNMKPEEYAHKILELQVQSILEKKMKVERSIAKQYNDNPS  120

Qy    902  LITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEII  961
           ||| |||||| :|:|| ||||||||||||||||||| :||||||||||| |:|||||||||
Db    121  LITLLCKNCSMLVCSGENIHVIEKMHHVNMTPEFKGLYIVRENKALQKKFADYQTNGEII  180

Qy    962  CKCGQAWGTMMVHKGLDLPCLKIRNFVVVFKNNSTKKQYKKWVELPITFPNLDYSECCLF 1021
           |||||||||||||||||||||||||||||:::||||||||||||||| :||||:|:||||
Db    181  CKCGQAWGTMMVHKGLDLPCLKIRNFVVNFKNNSPKKQYKKWVELPIRFPDLLDYSEYCLY  240

Qy   1022  SDED 1025
           ||||
Db    241  SDED  244
```

RESULT 14
O95786

```
ID   O95786;                PRELIMINARY;    PRT;    925 AA.
```

```
AC  O95786;
DT  01-MAY-1999 (TrEMBLrel. 10, Created)
DT  01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  RNA helicase.
GN  RIG-I.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Sun Y.W.;
RT  "RIG-I, a human homolog gene of RNA helicase, is induced by retinoic
RT  acid during the differentiation of acute promyelocytic leukemia
RT  cell.";
RL  Thesis (1997), Shanghai Institute of Hematology, Rui-Jin Hospital,
RL  Shanghai Second Medical University.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Yi-Wu S.;
RL  Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF038963; AAD19826.1; -.
DR  GO; GO:0005524; F:ATP binding; IEA.
DR  GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR  GO; GO:0016787; F:hydrolase activity; IEA.
DR  GO; GO:0003676; F:nucleic acid binding; IEA.
DR  GO; GO:0000166; F:nucleotide binding; IEA.
DR  InterPro; IPR003593; AAA_ATPase.
DR  InterPro; IPR001410; DEAD.
DR  InterPro; IPR001650; Helicase_C.
DR  Pfam; PF00270; DEAD; 1.
DR  Pfam; PF00271; helicase_C; 1.
DR  SMART; SM00382; AAA; 1.
DR  SMART; SM00487; DEXDc; 1.
DR  SMART; SM00490; HELICc; 1.
KW  ATP-binding; Helicase; Hydrolase.
SQ  SEQUENCE  925 AA;  106613 MW;  5922B4F3DDDF00D4 CRC64;

Query Match             20.5%;  Score 1086.5;  DB 4;  Length 925;
Best Local Similarity   31.2%;  Pred. No. 2.4e-52;
Matches  325;  Conservative 166;  Mismatches 364;  Indels 185;  Gaps 35;

Qy    12 RYLISCFRARVKMYIQVEPVLDYLT--FLPAEVKEQIQRTVATSGNMQAVELLLSTLEKG 69
Db     6 RRSLQAFQDYIRKTLDPTYILSYMAPWFREEEV-QYIQAEKNNKGPMEAATLFLKFLLE- 63

Qy    70 VWHLGWTREFVEALRRTGSPLAARYMNPELTDLSPSFENAH-DEYLQLLNLLIQPTLVD 127
Db    64 LQEEGWFRGFLDALDHAG--YSGLYEAIESWD-----FKKIEKLEEYRLLLKRIQPEFKT 116

Qy   128 KLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRIVQ--KENWFSAFLNVL 185
Db   117 RIIPTDIISD-LSECLLINQBCERILQICSTKGMMAGAEKLVECLLRSDKENWPKTL--- 171

Qy   186 RQTGNNELVQELTGSDCSESNAEIENLSQVDGPQVEEQLLSTTVQPNLEKEVWGMEN--- 242
Db   172 ------KLALEKERNKFSELWIVEKGIKDVETEDLEDKMETSDIQ------IFYQEDPEC 219

Qy   243 -NSSESSFADSSVVSESDTSLAEGSVSCLDESLGHNSNMGSDSGTMGSDSDEENVAARAS 301
Db   220 QNLSENSCPPSEV---SDTNL-------------------------------------YS 239

Qy   302 PEPELQLRPYQMEVAQPALEGKNIIICLPFGSGKTRVAVYIAKDHLDKKKKASBPGKVIV 361
Db   240 P--FKPRNVQLELALPAMKGKNTIICAPFGCGKTFVSLLICEHHL-KFPQGQGKGKVVF 295

Qy   362 LVNKVLLVEQ--LFRKEFQPFLKKWYRVIGLSGDTQLKISFPEVVKSCDIIISTAQIIE 418
Db   296 FANQIPVFPEQQKSVFSKYFE---RHGYRVTGISGATAENVPEQIVENNDIIILTPQILV 352

Qy   419 NSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPV 478
Db   353 N---NLKKGTIP--SLSIFTLMIFDECHNTSKQHPYNMIMFNYLDQKLGGSS------- 399

Qy   479 IPLPQILGLTASPGVGGATKQAAAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKK 538
Db   400 GPLPQVIGLTASVGVGDAKNTDEALDYICKLCASLDASVIATVKHNLEELEQVVYKPQKF 459

Qy   539 FAIADATREDPFKE-----------KLLEIMTRIQTYCQMSPMSDFGTQPYEQW 581
Db   460 FRKVESRISDKFKYIIAQLMRDTESLAKRICKDLENLSQIQN------REBFGTQKYEQW 512

Qy   582 AIQMEKKA----AKKGNRKERVC------AEHLRRYNEALQINDTIRMIDAYTHLETFYN 631
Db   513 IVTVQKACMVFQMPDKDEESRICKALFLYTSHLRKYNDALIISEHARMKDALDYLKDFFS 572

Qy   632 EEKDKKFAVIEDDSDEGGDEYCDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKRLA 691
Db   573 NVRAAGFEEIEQD----------------LTQRFEEKLQELESVS 601

Qy   692 ENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEK--FAEVGV 749
Db   602 RDPSNENPKLEDLCFILQEYHLNPETIT-ILFVKTRALVDALKNWIEGNPKLSFLKPGI 660

Qy   750 KAHHLIGAGHSSEFKPMTQNEQKEVISKFR-TGKINLLIATTVAEEGLDIKEBCNIVIRYG 808
Db   661 ----LTGRGKTNQNTGMTLPAQKCILDAFKASGDHNILIATSVADEGIDIAQCNLVILYE 716

Qy   809 LVTNEIAMVQARGRARADESTYVIVAHSGSGVIEHETVNDFREKMYKAIHCVQNMKPEE 868
Db   717 YVGNVIKMIQTRGRARGSKCFLLT-SNAGVIEKEQINMYKEKOMNNDSIIRLQTWDEAV 775

Qy   869 YAHKILELQMQSIMEKKMKTKRNIAKHY--KNNPSLITFLCKNCSVLACSGBEDIHVIEKM 926
Db   776 FREKLHIQTH---EKFIRDSQEKPKPVPDKENKKL---LCRKCKALACYTADVRVIEEC 829

Qy   927 HHVNMTPEFKELYIVRENKALQKKCADYQINGEIIC---KCGQAWGTMVHKGLDLPCLK 983
Db   830 HYTVLGDAFKECFVSRPHPX-PKQPSSFEKRAKIFCARQNCSHDWGIHVXKTFEIPVIK 888

Qy   984 IRNFVVFKNNSTKKQYKKW 1003
Db   889 IESFVVEDIATGVQTLYSKN 908
```

```
RESULT 15
ID   Q86X56
AC   Q86X56;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to melanoma differentiation associated protein-5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC046208; AAH46208.1; -.
SQ   SEQUENCE  221 AA;  25129 MW;  3C7139BCAC564BB5 CRC64;

Query Match             20.2%;  Score 1071;  DB 4;  Length 221;
Best Local Similarity   97.2%;  Pred. No. 2.6e-52;
Matches  210;  Conservative  2;  Mismatches  4;  Indels  0;  Gaps  0;

Qy     1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60
Db     1 MSNGYSTDENFRYLISCFRARVKMYIQVEPVLDYLTFLPAEVKEQIQRTVATSGNMQAVE 60

Qy    61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120
```

```
Db       61 LLLSTLEKGVWHLGWTREFVEALRRTGSPLAARYMNPELTDLPSPSFENAHDEYLQLLNL 120

Qy      121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRITVQKENWFSA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 LQPTLVDKLLVRDVLDKCMEEELLTIEDRNRIAAAENNGNESGVRELLKRITVQKENWFSA 180

Qy      181 FLNVLRQTGNNELVQELTGSDCSESNAEIENLSQVD 216
            ||||||||||||||||||||||||||| | ::|
Db      181 FLNVLRQTGNNELVQELTGSDCSESNAGICNFTEED 216
```

Search completed: June 2, 2004, 19:11:47
Job time : 53 secs